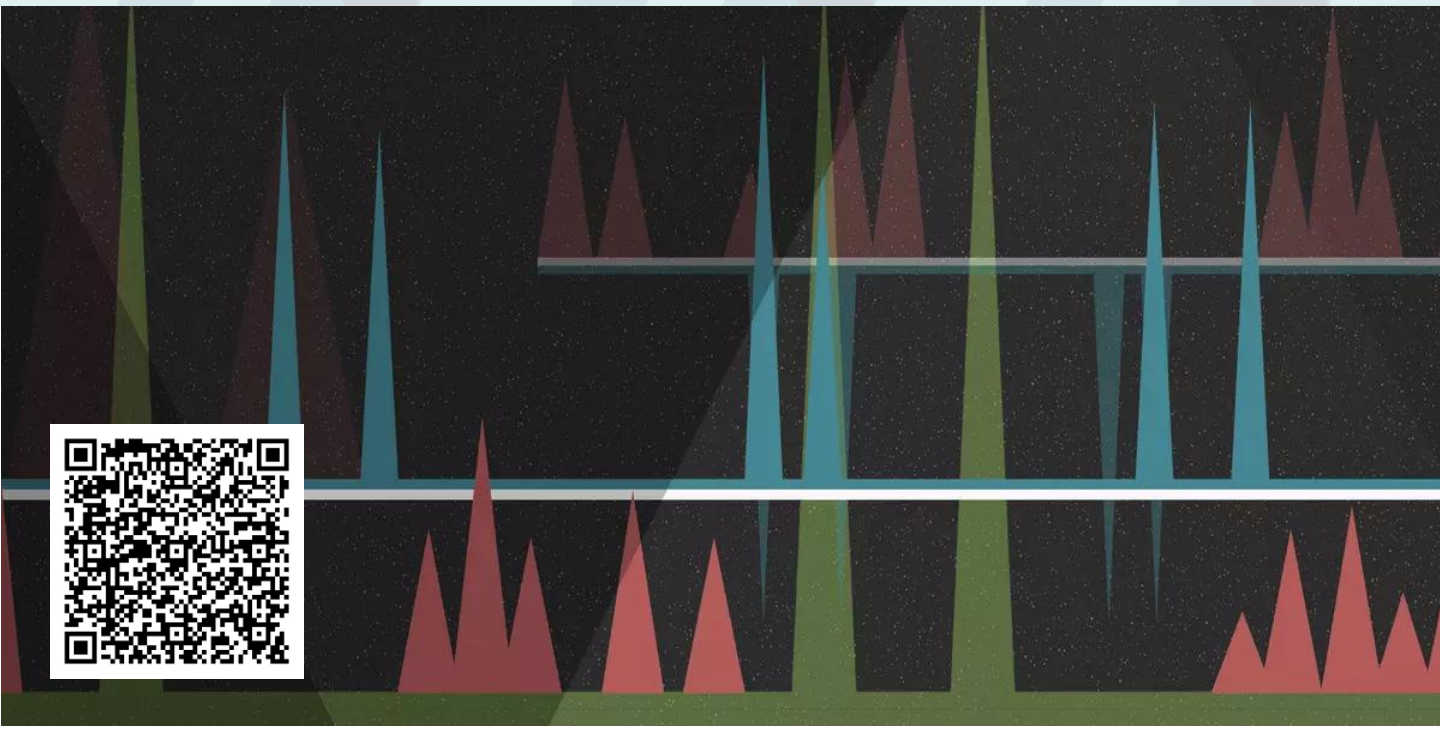
NIST IR 8351sup2
December 2024

Summarized Information from Publicly Accessible Validation Studies, Proficiency Test Results, and Interlaboratory Comparison Data

Supplemental Document to
DNA Mixture Interpretation:
A NIST Scientific Foundation Review



NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

These opinions, recommendations, findings, and conclusions do not necessarily reflect the views or policies of NIST or the United States Government. This report and its supplemental documents are not intended to imply any criticism of any past, pending, or future legal proceedings involving forensic science or evidence based on forensic science.

NIST Interagency Report
NIST IR 8351sup2

Summarized Information from Publicly Accessible Validation Studies, Proficiency Test Results, and Interlaboratory Comparison Data

Supplemental Document to DNA Mixture Interpretation: A NIST Scientific Foundation Review

John M. Butler

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8351sup2

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Interagency Report
NIST IR 8351sup2

Summarized Information from Publicly Accessible Validation Studies, Proficiency Test Results, and Interlaboratory Comparison Data

Supplemental Document to DNA Mixture Interpretation: A NIST Scientific Foundation Review

John M. Butler
Special Programs Office
Laboratory Programs
National Institute of Standards and Technology

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8351sup2

December 2024

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
Laurie E. Locascio, NIST Director and Under Secretary of Commerce for Standards and Technology

Certain equipment, instruments, software, or materials, commercial or non-commercial, are identified in this paper in order to specify the experimental procedure adequately. Such identification does not imply recommendation or endorsement of any product or service by NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

These opinions, recommendations, findings, and conclusions do not necessarily reflect the views or policies of NIST or the United States Government. This report and its supplemental documents are not intended to imply any criticism of any past, pending, or future legal proceedings involving forensic science or evidence based on forensic science.

NIST Technical Series Policies
Copyright, Use, and Licensing Statements
NIST Technical Series Publication Identifier Syntax

Publication History
Approved by the NIST Editorial Review Board on 2024-12-17

How to Cite this NIST Technical Series Publication
Butler, J.M. (2024) Summarized Information from Publicly Accessible Validation Studies, Proficiency Testing Results, and Interlaboratory Comparison Data: *Supplemental Document to DNA Mixture Interpretation: A NIST Scientific Foundation Review* (National Institute of Standards and Technology, Gaithersburg, MD), NIST IR 8351sup2. https://doi.org/10.6028/NIST.IR.8351sup2

Author ORCID iDs
John M. Butler: 0000-0001-6472-9157

Contact Information
john.butler@nist.gov

Preface

Much of the information in this supplemental document was originally part of the draft report within Chapter 4 when it was released for public comment on June 9, 2021. While preparing the final version of NISTIR 8351 **DNA Mixture Interpretation: *A NIST Scientific Foundation Review***, the authors decided to move this material into this supplemental document that will be available on the same NIST website[1].

Acknowledgments

This material was developed with valuable input from Sarah Riman, Peter Vallone, Hari Iyer, Melissa Taylor, and Sheila Willis. Additional edits and suggestions were provided by Laura Baxter, Erica Romsos, Robert Ramotowski, Kelly Sauerwein, and Katherine Sharpless. In addition, feedback from public comments provided on the draft report strengthened the final

[1] See https://www.nist.gov/spo/forensic-science-program/dna-mixture-interpretation-nist-scientific-foundation-review

content. The time and thought that went into each of the comments and suggestions is gratefully appreciated and acknowledged. Appreciation is expressed to those who have published or made their validation data or proficiency test results publicly accessible, especially Collaborative Testing Services, Inc. Elements of the cover art were created by Natasha Hanacek from the NIST Web Content Development Group.

Abstract

This report provides supplemental information to NISTIR 8351 **DNA Mixture Interpretation: *A NIST Scientific Foundation Review*** and summarizes information from publicly accessible validation and interlaboratory data as well as proficiency test results covering DNA mixture interpretation. The summary tables contain information extracted by the author from the sources listed and reflect coverage of various aspects and factors important to DNA mixture interpretation.

Keywords

DNA; DNA mixture interpretation; forensic genetics; forensic science; proficiency test results; scientific foundation review; validation data.

Table of Contents

List of Tables

List of Figures

1. **Introduction**

Supplemental documents are created with NIST scientific foundation reviews to provide additional information in support of key messages in the main report. This supplemental document should be used in conjunction with its accompanying report.

NISTIR 8351-draft was 250 pages in length when it was released for public comment in June 2021, in part because of long tables in Chapter 4 that summarized publicly accessible information. In addition to the summary tables, there are many accompanying references associated with these data sources. While preparing the final version of *DNA Mixture Interpretation: A NIST Scientific Foundation Review* (NISTIR 8351), the authors decided to move much of the summarized publicly accessible information into this supplemental document to retain important details for interested readers while not detracting from key messages shared in the main report. Table 4.4 from NISTIR 8351-draft, which summarizes studies involving comparison of results from multiple probabilistic genotyping software systems, has been retained in Chapter 4 of NISTIR 8351, and is now Table 4.2.

Each section below contains introductory remarks explaining the sources of data and how information was summarized in the accompanying tables. Following these informational tables, observations and comments are provided about the summarized data. Additional discussion regarding the significance of the findings and key takeaways are included in Chapter 4 of the main report (NISTIR 8351). New information has been added to each section since the draft report was prepared in 2021, including feedback received during the public comment period (e.g., PC12, PC38, and PC63, see NISTIR 8351-draft PCs) or new publications.

This document provides a summary of what has been observed and can be gleaned from publicly accessible data regarding DNA mixture interpretation as of 2024. Tables contain summarized information extracted by the author from the sources listed and reflect coverage of various aspects and factors important to DNA mixture interpretation.

Publicly accessible information on DNA mixture interpretation performance was examined from five types of sources: (1) published developmental validation studies from STR typing kits, (2) published PGS studies, (3) accessible PGS internal validation studies or summaries from forensic laboratories, (4) proficiency test results, and (5) published interlaboratory studies.

2. **Publicly Accessible Information**

Quality science is founded on empirical data. As previously written (NISTIR 8225), foundational scientific data should be *publicly accessible for independent review* so that interested parties can judge for themselves the value of the underpinning information. Others have also emphasized the importance of publicly accessible validation data sets, including multiple publications since the NISTIR 8351-draft report was released for public comment in June 2021 (e.g., GAO 2021, Chin & Ibaviosa 2022, Edge & Matthews 2022, Marciano & Maynard 2023, Kayser et al. 2023, Albright & Scurich 2024, EWG 2024). Researchers and practitioners are encouraged to share validation data[2] that informs the validity and reliability of forensic methods and practices and to make as much of this information publicly accessible as is permitted by law and human subjects research protections (e.g., Budowle & Sajantila 2023, Chapman et al. 2023, Marciano et al. 2023).

The FBI Quality Assurance Standards (QAS) and other accreditation requirements have not previously required validation data or summaries to be publicly accessible beyond "peer-reviewed publication of the underlying scientific principle(s) of a method"[3] involved in developmental validation. Forensic laboratories conduct internal validation studies to demonstrate that specific methods perform as expected in their individual environments – and thus the levels of performance deemed appropriate may vary across laboratories. Auditors do assess the types of studies performed as part of QAS audit procedures. Again, public accessibility of this information has not been a primary goal for most forensic laboratories.

Chapter 3 of NISTIR 8351 describes data sources explored in this scientific foundation review and strategies to locate information from validation experiments, proficiency tests, and interlaboratory comparison studies. Hundreds of articles on DNA mixture interpretation were collected from peer-reviewed journals, and many of them are cited throughout the main report (NISTIR 8351) and its accompanying supplemental documents (in addition to this one, see NISTIR 8351sup1).

Because this review considers DNA mixture interpretation using short tandem repeat (STR) markers, attention is given to any source of publicly accessible data involving DNA mixtures and not just performance involving probabilistic genotyping software (PGS) systems. With each of these sources, questions that may be answered using available data are considered. However, as noted in the preface for the final report (NISTIR 8351, pp. 7-8), the NIST authors of this study did not attempt to perform an assessment of global reliability of DNA mixture interpretation with publicly accessible data. Thus, this report and its supplemental documents are not intended for making any statements about the degree of reliability regarding information in any individual case in the past, present, or future.

---

[2] A July 2021 report from the Government Accountability Office noted (GAO 2021, p. 46): "The public may be more inclined to trust algorithms if officials provide free and easy access to results of operational testing, and to information about data sources, how algorithms are used, and for what types of investigations. For example, making operational testing results public could increase confidence in the accuracy and fairness of the algorithms."

[3] See Standard 8.2.2 in 2020 version of FBI Quality Assurance Standards for Forensic DNA Testing Laboratories; available at https://www.swgdam.org/publications (accessed October 28, 2024).

3. **Published Developmental Validation Studies for STR Typing Kits**

Validation studies and underlying experiments assist in assessing and understanding the degree of reliability of scientific methods and practices and are a key part of quality assurance efforts (Butler 2011, pp. 167-211). These studies are intended to help evaluate the degree to which a method is *robust* (i.e., successfully provides results a high percentage of the time), *reproducible* (i.e., provides the same or similar results when tested multiple times), and *reliable* (i.e., provides an accurate result that correctly reflects the sample being examined). When results from a method have been shown to be robust, reproducible, and reliable, then users of these results can have increased confidence that this method works appropriately on casework samples that possess characteristics similar to those of samples examined in the validation studies. The FBI QAS emphasize that validation studies should cover the types of DNA mixtures expected to be seen in casework: "Mixture interpretation validation studies shall include samples with a range of the number of contributors, template amounts, and mixture ratios expected to be interpreted in casework" (QAS 2020, 8.3.2.1). This range of characteristics has been termed *factor space coverage* in the accompanying report (NISTIR 8351).

Validation studies performed in a research environment or a practitioner laboratory provide information that allows the researchers or laboratory analysts and their stakeholders to make assessments regarding the degree of reliability for a particular method. Validation studies are designed to generate sufficient data such that the laboratory decision maker (e.g., DNA Technical Leader) can evaluate and decide whether a method is reliable for their application. Guidance documents on validation in forensic science typically focus on types of tests to perform in gathering data. Individual laboratories make decisions on how to assess the data or the number of samples needed to demonstrate a particular level of performance.

The FBI QAS and guidelines from the Scientific Working Group on DNA Analysis Methods (SWGDAM) have historically provided requirements and guidance on what types of studies to perform but have offered little guidance on how to determine the level of desired performance. A review of the language in these guidance documents is provided in the accompanying supplemental document covering the history of DNA mixture interpretation (see NISTIR 8351sup1, Table S1.4). For the forensic DNA community, levels of validation have been divided into (1) *developmental validation*, often performed under the auspices of the method developer, and (2) *internal validation*, performed within each user laboratory or laboratory system before employing a method for casework.

Results from developmental validation studies are more likely to be published in the peer-reviewed literature compared to internal validation studies. The secondary internal validation studies may not be viewed as novel enough for many scientific journals as has been previously noted (Buckleton 2009), yet their accessibility for independent assessments of reliability is vital.

Table S2.1 summarizes findings from three published developmental validation studies performed by the vendors of different commercial STR kits: GlobalFiler (Ludeman et al. 2018), Investigator 24plex (Kraemer et al. 2017), and PowerPlex Fusion 6C (Ensenberger et al. 2016). These kits are commonly used in U.S. forensic laboratories (see Table 3 in Brinkac et al. 2023). The findings and measurement experiments described in Table S2.1 for these three studies are

representative of other STR kit developmental validation studies (e.g., Krenke et al. 2002, Collins et al. 2004, Ensenberger et al. 2010, Wang et al. 2012, Green et al. 2013, Ensenberger et al. 2014, Oostdik et al. 2014).

Table S2.1. Summary of details and findings for measurement experiments and DNA mixture studies from three developmental validation studies of commonly used commercial STR typing kits. Abbreviations: SD = standard deviation; RFU = relative fluorescence units; nt = nucleotide.

#	Reference STR Kit (PCR Cycle #) Instruments Used	Measurement Experiments and Findings	Summary of Details and Findings on DNA Mixture Studies
1	Ludeman et al. (2018) **GlobalFiler** (29 cycles) ABI 3130xl, 3500, 3500xL	*Sensitivity*: Tested a single sample (007) from 3000 pg to 15.6 pg; found full profiles at ≥125 pg across 4 replicates; no significant saturation at 3 ng *Sizing precision*: Not reported *Reproducibility*: (see concordance) *Concordance*: Consistent genotypes with 1194 population samples against Identifiler and NGM SElect kits *Heterozygote balance*: Average ratios >80% (with 1 ng input DNA) *Stutter*: From 1092 population samples (table 4 in article); used mean + 3 SD	**Tested a single two-person mixture** (Raji & 007); genotypes were provided (28 of 43 alleles in 007 were non-overlapping); 1 ng total DNA used for all mixtures; 3 mixture ratios examined (1:1, 1:5, 1:8) and run in triplicate; detected all non-overlapping minor contributor alleles at the 1:5 ratio (167 pg minor) in six runs and in three of six runs at the 1:8 ratio (111 pg minor) using a 150 RFU analytical threshold
2	Kraemer et al. (2017) **Investigator 24plex QS & Investigator 24plex GO!** (30 cycles) ABI 3500, 3130	*Sensitivity*: Tested a single sample (9948) from 1000 pg to 8 pg; found full profiles consistently at ≥125 pg; for 8 pg, 50% of expected alleles were detected; no saturation at 1 ng *Sizing precision*: Sized alleles in 96 allelic ladders (max SD ≤0.08 nt) *Reproducibility*: Consistent genotypes in a single control DNA sample across three sites, eight replicates, two types of instruments *Concordance:* No null alleles from 656 NIST samples (99.997% with 29,520 alleles compared against six other STR kits) *Heterozygote balance*: Decreased towards lower template amounts (see fig. 10) *Stutter*: From 656 NIST population samples (table 1 in article); used max %	**Tested a single two-person mixture** (9948 & XX107); no genotypes or degree of allele overlap described; 500 pg total DNA used for all mixtures; nine mixture ratios examined (1:15, 1:10, 1:7, 1:3, 1:1, 3:1, 7:1, 10:1, 15:1,) and run in replicates of four; 100% of expected alleles were identified for minor components of 3:1, 7:1, and 10:1 mixtures; 97% of minor component alleles for 15:1 (31 pg minor) were identified using a 50 RFU analytical threshold

#	Reference STR Kit (PCR Cycle #) Instruments Used	Measurement Experiments and Findings	Summary of Details and Findings on DNA Mixture Studies
3	Ensenberger et al. (2016) **PowerPlex Fusion6C** (29 cycles) ABI 3130, 3130xl, 3500, 3500xL *Results from 8 laboratories*	*Sensitivity*: Tested in seven laboratories (seven 3500s, two 3130s) two DNA samples serially diluted from 2 ng to 31.25 pg with each amount run in replicates of four; with ABI 3500s, 99.7% of expected alleles were detected at 125 pg, 82% alleles at 62.5 pg, and 44% alleles at 31.25 pg; saturation at 2 ng on 3130s *Sizing precision*: Sized alleles from two injections of allelic ladders (8 to 48 depending on instrument; max SD ≤0.1 nt) *Reproducibility*: Concordant genotypes across six laboratories with NIST SRM 2391c and 2800M control DNA *Concordance:* Two discordant calls from 652 NIST samples (99.994% concordance in 33,558 alleles compared) *Heterozygote balance*: Not reported *Stutter*: From 652 samples (table 7 in article); used average + 1 SD	**Tested a single two-person mixture** in three laboratories; no genotypes or degree of allele overlap described; 1 ng total DNA used for all mixtures; 9 mixture ratios examined (1:19, 1:9, 1:5, 1:2, 1:1, 2:1, 5:1, 9:1, 19:1) in replicates of four; detected all non-overlapping minor contributor alleles at the 1:2 ratio (333 pg minor), 99% at 1:5 ratio (167 pg minor), 96% at 1:9 ratio (100 pg minor), and 74% at 1:19 ratio (50 pg minor) using analytical thresholds of 175 RFU for the 3500s and 50 RFU for the 3130s

3.1. **Comments on Published Developmental Validation Studies**

Developmental validation studies for STR typing kits typically focus on measurement aspects important for reliable genotyping of single-source DNA samples and parameters that can inform mixture interpretation guidelines, such as detection sensitivity, heterozygote balance (peak height ratios), and stutter ratios. These studies seek to answer questions about the robustness, reproducibility, and reliability of results using single-source DNA samples and exploring sensitivity performance by serially diluting samples from a few nanograms (ng) down to a few picograms[4] (pg) with each amount examined multiple times to assess reproducibility. Loss of expected alleles from a complete STR profile commonly begins to occur around 100 pg to 125 pg, or about 15 to 20 cell equivalents of DNA, due to stochastic effects when performing the polymerase chain reaction (PCR) with a limited number of starting DNA template molecules (Walsh et al. 1992).

An important focus of STR typing kit developmental validation studies involves *measurement capabilities* to demonstrate consistent and accurate allele calling of single-source samples using sizing precision studies, concordance with previous results, and reproducibility among multiple instruments or test sites. As conventional STR typing measures the length of PCR products using capillary electrophoresis, sizing precision of less than 0.5 nucleotides is important to distinguish

[4] A single diploid human cell contains approximately 6.6 pg so 1 ng of genomic DNA contains around 150 cells (Butler 2011, p. 52).

single nucleotide differences. Results from these types of studies have demonstrated a strong foundation in sizing precision and STR allele designation using allelic ladders and internal size standards with capillary electrophoresis measurements (e.g., Larazuk et al. 1998, Butler et al. 2004). Concordance with previous testing results from the same genetic marker demonstrates reproducibility and accuracy of repeated measurements[5] with new STR kits. Characterizing levels of heterozygote imbalance and quantity of stutter products that result from the PCR amplification of STR markers help inform a laboratory's mixture interpretation protocols.

The third column in Table S2.1 describes mixture studies performed in the three publications. Only a single two-person mixture combination was explored in each study with three to nine different mixture ratios, usually with replicate testing of each mixture ratio sample. The goal of these mixture studies is typically to demonstrate detection of non-overlapping alleles in minor contributors (i.e., *are measurement and interpretation possible?*) rather than accuracy in interpreting and/or deconvoluting mixture profiles (i.e., *are measurement and interpretation performed correctly?*).

Typically, for these developmental validation studies, limited information is collected involving DNA mixtures. It is rare for more than a single two-person mixture to be examined, with the mixture ratio being the primary variable explored. Overall success rate of detecting non-overlapping minor contributor STR alleles is a commonly used metric in these publications. Yet the degree of allele overlap, which depends on the genotype compositions of the mixture components, is not always described (e.g., rows 2 and 3 in Table S2.1). While these mixture studies address a requirement in guidance documents (e.g., see Table S1.4 in NISTIR 8351sup1), they offer limited information on performance of any DNA mixture interpretation protocols.

[5] "The power of DNA testing is only fully realized when results can be compared between laboratories in different areas or when offender samples present in DNA databases can be accurately matched with crime scene samples originating from that offender" (Butler 2011, p. 182).

4. Published PGS Studies

Published studies involving probabilistic genotyping software (PGS) were examined to learn more about DNA mixture interpretation performance in the last decade or so. There have been more than 70 articles published in the peer-reviewed literature with some form of PGS validation data (Table S2.2). Eight articles in this table (see bolded row number and reference entries) were examined and cited by the President's Council of Advisors on Science and Technology (PCAST) in their September 2016 report (PCAST 2016). Thus, a great deal more information is now available to assess the use of PGS in DNA mixture interpretation. Data summarized in Table S2.2 illustrate the information that exists for the experiments reported in these publications.

The following information, which relates to the factors cited in Table 4.1 of the main report (NISTIR 8351), were considered for the examined articles: publication year, author and title, PGS system and version number, STR typing kit used to generate the DNA profiles, the measured variables, whether results from multiple PGS systems were compared, number of samples, number of contributors, number of replicates, whether known samples were used for ground truth, source of DNA, amount of DNA, mixture ratios, sample condition (e.g., degraded DNA), degree of allele sharing in tested samples, total number of different individual samples contributing to the sample sets, non-contributor data construction and population(s) explored, and whether likelihood ratios (LRs) data points were reported.

Only a portion of this information is displayed in Table S2.2 as many of the publications did not contain, or did not clearly describe, all the information sought for this review. What is provided here summarizes the aspects most common in the publications examined. Those studies involving comparisons across more than one PGS system are also summarized in the main report (Table 4.2 in NISTIR 8351).

This list is not exhaustive. There have been other publications examining specific topics like investigating the impact of related people on non-donor likelihood ratios and when siblings of the true mixture contributor may be examined (e.g., Kalafut et al. 2022a, Kalafut et al. 2022b, Kelly et al. 2022b), creating database search compatible profiles from a PGS system (e.g., Myers 2021), or using PGS with single-cell STR analysis (e.g., Huffman & Ballantyne 2022, Huffman & Ballantyne 2023).

Table S2.2. NIST-extracted information on published PGS studies from peer-reviewed literature grouped by PGS system and publication date. Studies listed on rows #6, #7, #10, #11, #12, #13, #14, and #57 were part of the PCAST 2016 review (row numbers and references are in bold font). When multiple values are included in a cell, a vertical line (|) is used to separate them and to correspond with information from adjacent column values (e.g., NoC and number of samples examined by NoC). NoC = number of contributors; N.E.S. = not explicitly stated in the referenced publication; N/A = not applicable; *comparison of multiple PGS systems are discussed in Table 4.2 of the main report (NISTIR 8351). †inclusion of ranges is not meant to imply that all combinations of DNA quantities and mixture ratios were covered. $ a 31-laboratory compilation (Bright et al. 2018) contained data from 8 different STR kits: GlobalFiler, Identifiler Plus, NGM SElect, PowerPlex Fusion 5C, PowerPlex Fusion 6C, PowerPlex ESI17 Pro, PowerPlex ESI17 Fast, and PowerPlex 16 HS.

#	Reference	PGS System *STR Kit*	NoC Examined	# samples by NoC	Total DNA Quantity Range (pg)†	Mixture Ratio Range†
1	Perlin & Sinelnikov 2009 (see also Perlin 2018)	TrueAllele *PowerPlex 16*	2	40	125 to 1000	1:1 to 9:1
2	Perlin et al. 2011	TrueAllele *Pro+Cofiler*	2	16 adjudicated cases	N.E.S.	N.E.S.
3	Perlin et al. 2013	TrueAllele *Pro+Cofiler*	2| 3	73| 14 adjudicated cases	N.E.S.	N.E.S.
4	Ballantyne et al. 2013 (proof of concept)	TrueAllele *Identifiler*	2	2	N.E.S.	1:1
5	Perlin et al. 2014	TrueAllele *PowerPlex 16*	2| 3| 4	40| 65| 8 adjudicated cases	N.E.S.	N.E.S.
6	**Perlin et al. 2015**	TrueAllele *Identifiler Plus*	2| 3| 4| 5	10| 10| 10| 10 (5 donors)	200, 1000	1:1 to 32:16:15:2:1
7	**Greenspoon et al. 2015**	TrueAllele *PowerPlex 16*	1| 2| 3| 4	11| 18| 15| 7 (11 donors)	10 to 1000	1:1 to 17:1:1:1
8	Bauer et al. 2020	TrueAllele *Fusion 5C*	2| 3| 4| 5| 6| 7| 8| 9| 10	2| 2| 2| 2| 2| 2| 2| 2| 2	500	2:1 to 25:19:14:13:12:6:5:3:1:1
9	Taylor et al. 2013	STRmix *Identifiler & NGM SElect*	Ex. 1: 2 Ex. 2: 2| 3	Ex. 1: 127 (ID) Ex. 2: 4| 6 (NGM)	100 to 500	1:1 to 5:1, 3p mixes (N.E.S.)
10	**Bright et al. 2014**	STRmix *Identifiler*	2| 3	1| 9	1500	1:1, 1:1:1, 10:5:1
11	**Taylor 2014**	STRmix *GlobalFiler*	2| 3| 4	15| 6| 10 (4 donors)	10 to 400	1:1 to 10:1; 1:1:1 to 4:3:2:1
12	**Taylor & Buckleton 2015**	STRmix *GlobalFiler*	4	29 profiles (Taylor 2014 data)	10 to 400	1:1:1:1 or 4:3:2:1
13	**Taylor et al. 2015**	STRmix *GlobalFiler & Profiler Plus*	1| 2| 3| 4	4| 1| 1| 3 (3 GlobalFiler & 6 Profiler Plus tests)	10 to 500	1:1 to 4:3:2:1
14	**Bright et al. 2016**	STRmix *GlobalFiler*	2| 3| 4	93 profiles (Taylor 2014 data)	10 to 400	1:1 to 10:1; 1:1:1 to 4:3:2:1
15	Taylor et al. 2016a	STRmix *GlobalFiler*	N.E.S.	205 profiles	N.E.S.	N.E.S.

#	Reference	PGS System *STR Kit*	NoC Examined	# samples by NoC	Total DNA Quantity Range (pg)†	Mixture Ratio Range†
16	Taylor et al. 2016b	STRmix *6 different kits*	N.E.S.	1867 profiles in 14 datasets	N.E.S.	N.E.S.
17	Taylor et al. 2017a	STRmix *multiple kits*	1| 2	N.E.S.	N.E.S.	N.E.S.
18	Taylor et al. 2017b	STRmix *GlobalFiler*	4	29 profiles (Taylor 2014 data)	10 to 400	1:1:1:1 or 4:3:2:1
19	Taylor et al. 2017c	STRmix *GlobalFiler & Profiler Plus*	1| 2| 3	1| 3| 1	50 to 1000	1:1 to 10:1; 3:2:1
20	Moretti et al. 2017	STRmix *Identifiler*	1| 2| 3| 4| 5	>1400| 105| 64| 84| 24	19 to 4000 (their table 1)	1:1 to 10:1:1:2:2 (their table 1)
21	Bright et al. 2018 *(combined data from 31 labs)*	STRmix *8 different kits*§	3| 4| 5| 6	1315| 1263| 182| 65 (combined data)	N.E.S.	N.E.S.
22	Kelly et al. 2018	STRmix *GlobalFiler*	2| 3	35| 36 (PROVEDIt data)	6 to 750	1:1, 4:1, 9:1; 1:1:1, 1:4:1, 4:4:1
23	Bille et al. 2019	STRmix *GlobalFiler*	3| 4| 5	24| 73| 50 (60 mixtures, 147 interpretations)	250 to 1000	98:1:1 to 75:20:2:2:1
24	Bright et al. 2019b	STRmix *GlobalFiler*	2| 3| 4| 5	6| 6| 6| 6 (PROVEDIt data)	126 to 750 (their table 1)	1:1 to 1:9:9:9:1 (their table 1)
25	Noël et al. 2019	STRmix *Identifiler Plus*	4	24 = 12 known + 12 casework	160 to 3260	1:1:1:1 to 10:5:2:1
26	Duke & Myers 2020	STRmix *GlobalFiler*	1| 2| 3| 4	1| 2| 4| 4 (4 donors)	250 to 1000 (degraded DNA)	1:1 to 7:1:1:1
27	Lin et al. 2020	STRmix *GlobalFiler*	3	40 profiles tested (3 related donors)	100 to 500	1:1:1 to 10:10:1
28	Schuerman et al. 2020	STRmix *GlobalFiler*	3| 4	26| 33	100 to 1000	1:1:1 to 1:1:1:1 to 20:4:4:1
29	McGovern et al. 2020	STRmix *Fusion 5C*	2| 3| 4	Ex. 1: 2| 3| 5 Ex. 2: 11| 10| 10	150 to 1500	1:1 to 20:1; 5:1:1:1 to 10:5:5:1
30	Ward et al. 2022	STRmix *GlobalFiler RapidHIT ID*	2| 3| 4	6| 8| 10 (DNA) 6| 6| 0 (fluid mix) 29 (touch DNA)	(DNA) 50000 put on each of two swabs	1:1 to 1:10; … 1:1:1:1 to 1:5:10:10
31	Duke et al. 2022	STRmix *GlobalFiler*	2| 3| 4	49| 74| 74 (with multiple levels of allele sharing)	25 to 3600 (with many different combinations)	1:1 to 99:1 … 1:1:1:1 to 100:100:100:6
32	Duke et al. 2023	STRmix *GlobalFiler*	1| 2| 3| 4	26| 49| 74| 74 (with multiple levels of allele sharing)	25 to 3600 (with many different combinations)	1:1 to 99:1 … 1:1:1:1 to 100:100:100:6
33	Riman et al. 2024a	STRmix *GlobalFiler*	1| 2| 3| 4| 5| 6	19| 59| 57| 55| 65| 10	8 to 4000	265 combinations (e.g., 1:1 to 1:9:19:29:38:5)

#	Reference	PGS System *STR Kit*	NoC Examined	# samples by NoC	Total DNA Quantity Range (pg)†	Mixture Ratio Range†
34	Benschop et al. 2019a	EuroForMix *Fusion 6C*	2| 3| 4| 5	30| 30| 30| 30	180 to 900	1:1 to 20:1:2:1:1
35	Benschop et al. 2019b	DNAxs *Fusion 6C*	1| 2| 3| 4	20| 10| 10| 10 (simulated profiles)	N/A (simulated data)	N/A (simulated data)
36	Benschop et al. 2020	DNAxs *Fusion 6C*	1| 2| 3| 4| 5	17| 38| 38| 37| 12 (71 donors)	180 to 5350	1:1 to 20:2:1:1:1
37	Lucassen et al. 2021	Mixture Solution *Identifiler Plus*	2	15 (11 tests, 4 casework)	tests: 1000 (casework: 20 to 60)	2:1 to 1:50
38	Holland et al. 2022	MaSTR *Fusion 6C*	2	144	100 to 500	1:1 to 1:10
39	Adamowicz et al. 2022	MaSTR *Fusion 5C*	2| 3| 4| 5	120| 96| 36| 36 (with low & high allele sharing)	70 to 2500	1:1 to 1:10; 1:1:1:1:1 to 1:2:2:5:10
40	Manabe et al. 2022	Kongoh *GlobalFiler*	2| 3| 4	40| 40| 40 (7| 12| 8 degraded PROVEDIt data)	250 to 1000 (45 to 750)	1:1 to 19:1; … 1:1:1:1 to 7:1:1:1
41	Gill & Haned 2013	LRmix *SGM Plus*	N.E.S.	3 examples with non-contributor performance tests	N.E.S.	N.E.S.
42	Benschop et al. 2012	LRmix *NGM*	2| 3| 4	1| 2| 1 (8 donors)	180 to 390	5:1 to 10:1:1:1
43	Benschop et al. 2015a	LRmix Studio *NGM*	1| 2| 3	64| 64| 64	3 to 36	1:1 to 1:1:1
44	Benschop et al. 2015b	LRmix *NGM*	3| 4| 5	12| 12| 12 (60 donors)	1250 to 1750	2:2:1 to 2:2:1:1:1
45	Haned et al. 2015	LRmix *NGM*	3| 4| 5	76| 74| 61	50 to 500	2:1:1 to 10:10:5:5:5
46	Haned et al. 2016	LRmix *NGM*	N.E.S.	77 mixtures; 1095 LRs	N.E.S.	N.E.S.
47	Kalafut et al. 2018	ArmedXpert *GlobalFiler*	1| 2| 3| 4	368| 64| 54 | 54 (67 donors)	100 to 1000	1:1 to 80:1; 1:1:1:1 to 20:4:4:1
48	Mitchell et al. 2012	FST *Identifiler*	1| 2| 3| 4	15| 214| 232| 31 (85 donors)	6.25 to 500	1:1, 4:1 1:1:1, 5:1:1
49	Balding 2013	likeLTD *Identifiler, SGM Plus*	1| 2| 3| 4	3| 5| 1| 1	N.E.S.	N.E.S.
50	Steele et al. 2014	likeLTD *SGM Plus*	1| 2| 3	3| 2| 4 (5 donors)	15 to 500	17:1 to 1:1:1
51	Steele et al. 2016	likeLTD *NGM SElect*	1| 2| 3	36| 24| 12 (36 donors)	4 to 328	1:1 to 16:1; 1:1:1 to 16:4:1
52	Puch-Solis et al. 2013	DNA Insight *SGM Plus*	1| 2	560 profiles (14 donors)	50 to 1500	1:1 to 9:1

#	Reference	PGS System *STR Kit*	NoC Examined	# samples by NoC	Total DNA Quantity Range (pg)†	Mixture Ratio Range†
53	Swaminathan et al. 2016	CEESIt *Identifiler Plus*	1| 2| 3	303 total	8 to 1000	1:1 to 49:1; 1:1:1 to 9:9:1
54	Benschop & Sijen 2014	LoCIM tool *NGM*	2| 3| 4	Training: 5| 13| 6 Testing: 70| 34| 27	60 to 1200	1:1 to 15:7:1:1
55	Bleka et al. 2019	CaseSolver *Fusion 6C*	2| 3| 4	9| 12| 4 (14 donors)	1000	1:1 to 13:1:1 to 4:4:1:1
56	Benschop et al. 2017b	SmartRank *NGM+SE33*	2| 3| 4| 5	155| 155| 16| 17	N.E.S.	N.E.S.
57	**Bille et al. 2014**	*multiple *Identifiler*	2	50 (2 donors)	100 to 500	1:1 to 5:1
58	Puch-Solis & Clayton 2014	*multiple *SGM Plus*	1| 2| 3| 4	10 replicates| 5| 1| 1 (Balding 2013 data)	N.E.S.	N.E.S.
59	Bright et al. 2015	*multiple *Identifiler*	2	Simulated profiles (2 donors)	N/A (simulated data)	1:1; 3:1
60	Bleka et al. 2016a	*multiple *ESX17*	1| 2| 3| 4	N.E.S.	N.E.S.	1:1 to 9:1; 5:4:1; 5:2:2:1
61	Bleka et al. 2016b	*multiple *NGM*	2| 3	4| 55 (33 donors)	180 to 1000	5:1 to 10:5:1
62	Manabe et al. 2017	*multiple *Identifiler Plus*	2| 3| 4	27| 27| 18	250 to 1000	1:1 to 7:1:1:1
63	Swaminathan et al. 2018	*multiple *Identifiler Plus*	1| 2| 3	30| 41| 30	16 to 1000	1:1 to 9:9:1
64	Alladio et al. 2018	*multiple *7 kits*	2| 3	3| 4	500 (1 diluted to 4)	1:1, 8:1, 19:1; 1:1:1 to 20:9:1
65	Buckleton et al. 2018	*multiple *Identifiler Plus*	2| 3| 4	2| 2| 1 (NIST MIX13 data)	N.E.S.	1:1 to 1:1:1:1
66	Rodriguez et al. 2019	*multiple *PowerPlex 21*	2	102	500	1:1 to 19:1
67	You & Balding 2019 (data from Steele et al. 2016)	*multiple *NGM SElect*	1| 2| 3	36| 24| 12 (36 donors)	4 to 328	1:1 to 16:1; 1:1:1 to 16:4:1
68	Riman et al. 2021 (the same data are discussed in Buckleton et al. 2022, Riman et al. 2022)	*multiple *GlobalFiler*	2| 3| 4	154| 147| 127 (PROVEDIt data)	30 to 750	1:1 to 1:9; … 1:1:1:1 to 1:9:9:1
69	Cheng et al. 2021 (Buckleton et al. 2024)	*multiple *GlobalFiler*	1| 2| 3| 4	8| 79| 30| 26 (PROVEDIt data)	30 to 750 (ss: 8 to 500)	1:1 to 1:9; 1:1:1:1 to 1:9:9:1
70	Susik & Sbalzarini 2023 (same as Riman et al. 2021)	*multiple *GlobalFiler*	2| 3| 4	154| 147| 127 (PROVEDIt data)	30 to 750	1:1 to 1:9; … 1:1:1:1 to 1:9:9:1
71	Greenspoon et al. 2024	*multiple *Fusion 5C*	2| 3| 4	19| 22| 12	500	1:1 to 19:1 1:1:1 to 30:2:1 1:1:1:1 to 9:3:2:1
72	McCarthy-Allen et al. 2024 (same as Riman et al. 2021)	*multiple *GlobalFiler*	2| 3| 4	154| 147| 127 (PROVEDIt data)	30 to 750	1:1 to 1:9; … 1:1:1:1 to 1:9:9:1

4.1. **Comments on Published PGS Studies**

Table S2.2 summarizes NIST-extracted information from 72 published articles on PGS and associated validation studies from the peer-reviewed literature. This compilation explored the numbers of samples and types of DNA mixtures examined in the reported publications. These studies are not necessarily focused on global reliability of the method used, and some of these articles focus on only a portion of the DNA mixture interpretation process. Interested readers can examine the original publications to learn more.

This compilation comes from at least 20 different PGS systems or precursors:

- STRmix (34 studies)
- LRmix or LRmix Studio (13 studies)
- EuroForMix (12 studies)
- TrueAllele (9 studies)
- likeLTD (5 studies)
- Lab Retriever (4 studies)
- DNAxs (2 studies)
- CEESIt (2 studies)
- Kongoh (2 studies)
- MaSTR (2 studies)
- Mixture Solution (2 studies)
- And one study each with ArmedXpert, FST, DNA Insight, LoCIM tool, CaseSolver, SmartRank, DNAmixtures, Hamilton Monte Carlo sampling, and DNAStatistX.

A variety of STR typing kits were used in combination with these various PGS systems. In addition, there are 20 studies comparing assigned LR values from multiple PGS systems that are summarized within the main report (NISTIR 8351, Table 4.2).

Occasionally details are not readily accessible or not explicitly stated (N.E.S.) in the publication; for example, the range of total DNA quantities or mixture ratios examined. Missing information points to areas where standardization of data reporting could potentially be developed for future journal publications on DNA mixture interpretation. A similar issue was observed in the 20 publicly accessible PGS internal validation summaries examined (see Table S2.4). The details listed in Table S2.2 were NIST-extracted and condensed in the format listed. This process was an objective attempt to extract pertinent information from the papers, but certain aspects or details that the authors intended to convey may not have been carried over.

4.1.1. General Metrics and Trends from These Publications

Forty-four of the 72 studies (61%) examined one or more four-person mixtures and 14 (19%) considered five-person mixtures. One publication examined up to 10-person mixtures (Bauer et al. 2020). Total DNA quantities explored ranged from 10 pg to 4000 pg with mixture ratios from 1:1 to 99:1 for two-person mixtures. With higher-order mixtures involving three or more contributors, various combinations and mixture ratios were utilized across these PGS studies to explore the assignment of LR values to minor contributors contributing sometimes <5% of the total DNA in the mixture. *Each study was designed to examine different aspects of PGS performance and therefore needs to be examined in the context of experiments performed and questions asked by the authors of those studies.* Some trends could be noted, but no definitive global reliability assessments could be made by examining these studies.

Some studies discuss allele sharing amongst contributors. For example, one study reported:

> "It appears [based on Fig. 10 in their publication] that the greater the allele sharing, the less the power there is to discriminate a true contributor from a non-contributor… However, further experimentation shows that this apparent trend is totally confounded by the number of contributors to the mixture." (Bright et al. 2018).

This comment came from the largest study (see row 21 in Table S2.2), which examined 2825 mixtures as combined data from 31 laboratories using 8 different STR kits and 2 capillary electrophoresis platforms (Bright et al. 2018). Participating laboratories provided the PGS developer with ground-truth known profiles coming from samples prepared as three-, four-, five-, or six-person mixtures as part of their own STRmix internal validation studies as well as reference profiles for the known contributors, their laboratory-specific settings, and the apparent number of contributors to each profile. A primary conclusion from this study stated:

> "In principle, we observe less discriminatory LRs for true and non-contributors when the APH [apparent peak heights for the DNA template] decreases per contributor. Again, this does not mean that mixed DNA profiles with contributors containing less DNA are unreliable, just they are less informative with respect to the true and non-contributors" (Bright et al. 2018).

4.1.2. What Data Would Be Helpful in Future Publications?

Given knowledge gained from available PGS studies, what data would be helpful in future publications? There are currently no documentary standards describing what types of information and details should be shared in published PGS studies to aid independent review of this data (see Chapter 4 and Box 4.1 in NISTIR 8351). Having and adhering to such standards would improve standardization.

A NIST research study published in September 2021 (Riman et al. 2021) examined two PGS systems using the publicly accessible PROVEDIt dataset (Alfonse et al. 2018; see also Table 3.2 in NISTIR 8351). This study provided examples of empirical assessment approaches using discrimination performance of H_1-true and H_2-true LR distributions and shared example outputs with assigned LR values as data points in supplemental files (see Riman et al. 2021, Tables S4

and S5). Availability of assigned LR values in this format enabled independent review and assessment by authors not associated with the original study (Buckleton et al. 2022, Riman et al. 2022). Other researchers have performed similar studies evaluating the same 428 mixture samples with different PGS systems (e.g., Susik & Sbalzarini 2023, McCarthy-Allen et al. 2024).

5. **Publicly Accessible PGS Internal Validation Studies**

During our discussions early in this project on the topic of available data to assess PGS systems for DNA mixture interpretation performance, the DNA Resource Group (see Table 1.2 in NISTIR 8351) underscored that additional PGS data exists in forensic laboratories as part of their internal validation studies. As described in Chapter 3, internet searches were conducted to locate publicly accessible internal validation data or information (see Table S2.3 for links to the summaries that could be found when these searches were performed). Table S2.4 summarizes information described in these validation studies and important details missing or not explicitly stated with the publicly accessible information.

Table S2.3. Publicly accessible PGS internal validation summaries from forensic laboratories located via multiple internet searches (initial search on March 23, 2020 with updates on February 8, 2021, and March 3, 2023). See Table S2.4 for details available from these validation summaries. Some information became available during the draft report public comment period. All of the websites listed below were accessible as of October 2024.

Laboratory	Information Available and Website
California Department of Justice DNA Laboratory (Richmond, CA)*	STRmix v2.06 (Identifiler Plus, ABI 3130/3500) https://epic.org/state-policy/foia/dna-software/EPIC-16-02-02-CalDOJ-FOIA-20160219-STRmix-V2.0.6-Validation-Summaries.pdf (45 pages)
Colorado Bureau of Investigation Laboratory System (Denver, CO)**	STRmix v2.5.11 (GlobalFiler, ABI 3500xL) https://indefenseof.us/uploads/Colorado-Bureau-Investigation-2018-STRmix-Validation_Summary.pdf (107 pages)
Erie County Central Police Services Forensic Laboratory (Buffalo, NY)	STRmix v2.3 (PowerPlex Fusion, ABI 3500) https://johnbuckleton.files.wordpress.com/2016/09/strmix-implementation-and-internal-validation-erie-fusion.pdf (32 pages) STRmix v2.3 (Identifiler Plus, ABI 3500) https://johnbuckleton.files.wordpress.com/2016/09/strmix-implementation-and-internal-validation-erie-id-plus.pdf (41 pages)
Idaho State Police Forensic Services Laboratory (Meridian, ID)**	STRmix v2.8 (PowerPlex Fusion 5C, ABI 3500) https://www.nist.gov/system/files/documents/2021/12/03/Public-Comments-NISTIR8351draft_0.pdf (46 pages; see pp. 401-446); for current summaries, see https://isp.idaho.gov/forensics/validation-summaries/
Jefferson County Regional Crime Laboratory (Golden, Colorado)**	STRmix v2.6 (GlobalFiler, ABI 3500) https://indefenseof.us/uploads/Jefferson-County-STRmix-Validation-V2.6-V2.6.3.pdf (78 pages)
Las Vegas Metropolitan Police Department (Las Vegas, NV)**	STRmix v2.6 (Investigator 24plex QS, ABI 3500xL) https://indefenseof.us/uploads/LVMPD-Summary.pdf (32 pages) and https://www.nist.gov/system/files/documents/2021/12/03/Public-Comments-NISTIR8351draft_0.pdf (see link on p. 47 for PC12 in NISTIR 8351-draft PCs)
Los Angeles County Sheriff's Department (Los Angeles, CA)**	STRmix v2.5.11 (PowerPlex Fusion 6C, ABI 3500) https://indefenseof.us/uploads/LASD-STRmix-2.5.11-Validation-Summary.pdf (84 pages)
Michigan State Police (Lansing, MI)	STRmix v2.3.07 (PowerPlex Fusion, ABI 3500/3500xl) https://johnbuckleton.files.wordpress.com/2016/09/strmix-summary-msp.pdf (47 pages); see also PC38 in NISTIR 8351-draft PCs
Office of Chief Medical Examiner Forensic Biology Laboratory (New York City, NY)	https://www1.nyc.gov/site/ocme/services/validation-summary.page STRmix v2.4 (PowerPlex Fusion, ABI 3130xl) (51 pages) STRmix v2.7 (PowerPlex Fusion 5C, ABI 3500xL) (82 pages + 34 pages on parameters)

Laboratory	Information Available and Website
Oregon State Police Forensic Services Division (Portland, OR)**	STRmix v2.3 and v2.4 (Identifiler Plus, GlobalFiler, ABI 3500xL) https://indefenseof.us/uploads/Oregon-State-Police-Portland-Metro-Lab-DNA-Val-067-GlobalFiler-STRmix-Summary_Redacted.pdf (42 pages)
Palm Beach County Sheriff's Office (West Palm Beach, FL)	STRmix v2.4 (PowerPlex Fusion, ABI 3500xl)** https://indefenseof.us/uploads/PBSO-STRmix-2.4-validation-summary.pdf (53 pages) STRmix v2.6.2 (PowerPlex Fusion 6C, ABI 3500xl) https://pbso.qualtraxcloud.com/showdocument.aspx?ID=10787 (45 pages)
Sacramento County District Attorney's Crime Laboratory (Sacramento, CA)**	STRmix v2.4 (PowerPlex Fusion 6C, ABI 3500xL) https://indefenseof.us/uploads/Sacramento-Cty-DA-STRmix-V2-4-internal-validation-summary.pdf (45 pages)
San Diego Police Department Crime Laboratory (San Diego, CA)	STRmix (GlobalFiler, ABI 3500), STRmix v2.3.06; STRmix v2.3.07; STRmix v2.4.06; STRmix v2.6 https://www.sandiego.gov/police/services/crime-laboratory-documents STRmix v2.3.07 (GlobalFiler, 5-person mixtures) (17 pages) https://www.sandiego.gov/sites/default/files/validationsummarystrmix-additionalstudiesmarch2016.pdf STRmix v2.4.06 Performance Check (16 pages) – *not included in Table S2.4* https://www.sandiego.gov/sites/default/files/strmix-performancecheckv2_4_06.pdf STRmix v2.6 Performance Check (39 pages) - *not included in Table S2.4* https://www.sandiego.gov/sites/default/files/strmixv2.6validationandperformancecheck_08-01-2019.pdf
Virginia Department of Forensic Science (Richmond, VA)*	TrueAllele Casework (PowerPlex 16, ABI 3130xl) https://epic.org/state-policy/foia/dna-software/EPIC-15-10-13-VA-FOIA-20151104-Production-Pt2.pdf (27 pages)
District of Columbia Department of Forensic Sciences (Washington, DC)	https://dfs.dc.gov/page/fbu-validation-studiesperformance-checks STRmix v2.3 (Identifiler Plus, ABI 3130) (50 pages + 15 pages on parameters) STRmix v2.4 (GlobalFiler, ABI 3500) (57 pages + 17 pages on parameters) STRmix v2.4 Supplemental Interpretation Study "The Zoom Study" (22 pages)
Wisconsin State Crime Laboratory (Milwaukee & Madison, WI)**	STRmix v2.6 (PowerPlex Fusion 6C, ABI 3500xL) https://indefenseof.us/uploads/Wisconsin-STRmix-Validation-Summary-Part-1-Single-Single-Source-to-Three-Person-Mixtures.pdf (38 pages)

*Information available online via a Freedom of Information Act request by the Electronic Privacy Information Center (epic.org)
** Information made available via Brooklyn Defender Services (https://indefenseof.us/issues/kinship-problem) after our draft report was released in June 2021 or as part of the public comment record (NISTIR 8351-draft PCs)

Table S2.4. NIST-extracted information from PGS internal validation studies listed in Table S2.3 and publicly accessible as of March 3, 2023. NoC = number of contributors; N.E.S. = not explicitly stated in the referenced public source; N/A = not applicable; F = female; M = male; *mixtures with related individuals; †inclusion of ranges is not meant to imply that all combinations of DNA quantities and mixture ratios were covered.

Laboratory PGS (version) *STR Kit* ABI CE	NoC Range	# samples	Total DNA Quantity Range (pg)†	Mixture Ratios Range†
California Department of Justice DNA Lab (Richmond, CA) STRmix (v2.0.6) *Identifiler Plus* ABI 3130 & 3500	1	N.E.S.	16, 31, 62, 125, 250, 500, 1000, 2000	N/A
	2	N.E.S.	500 | 1000	9:1, 4:1, 1:1 | 19:1, 9:1, 4:1, 2:1, 1:1
	3	N.E.S.	250, 375, 500, 750, 1000, 1500	1:1:1, 4.5:4.5:1, 6:3:1, 8:1:1
Colorado Bureau of Investigation Laboratory System (Denver, CO) STRmix (v2.5) *GlobalFiler* ABI 3500xL	1	N.E.S.	16, 31, 63, 125, 250, 500, 1000, 4000	N/A
	2	52 | 72*	250, 500, 1000, 1500, 2000	20:1, 10:1, 5:1, 3:1, 1:1
	3	42 | 96*	250, 500, 1000, 1500, 2000	10:5:1, 10:10:1, 3:2:1, 1:1:1
	4	52 | 48*	250, 500, 1000, 1500, 2000	4:3:2:1, 10:5:2:1, 10:10:5:1, 10:5:5:1, 1:1:1:1
	5	42	250, 500, 1000, 1500, 2000	5:4:3:2:1, 10:5:5:2:1, 10:10:5:2:1, 1:1:1:1:1
Erie County Forensic Laboratory (Buffalo, NY) STRmix (v2.3) *PowerPlex Fusion* 30 cycles ABI 3500	1	95	N.E.S.	N/A
	2	N.E.S.	500	19:1, 9:1, 3:1, 1:1
	3	N.E.S.	37, 75, 150, 300, 600 | 12, 25, 50, 100, 200, 400 | 500	3:2:1 | 1:1:1 | 5:1:1, 10:4:1, 1:5:1, 4:1:10, 1:1:5, 1:10:4
	4	N.E.S.	62, 125, 250, 500, 1000 | 500	4:3:2:1 | 17:1:1:1; 14:3:2:1; 1:1:1:1
Erie County Forensic Laboratory (Buffalo, NY) STRmix (v2.3) *Identifiler Plus* 29 cycles ABI 3500	1	94	N.E.S.	N/A
	2	N.E.S.	19, 37, 75, 150, 300 | 12, 25, 50, 100, 200, 400 | 500	2:1 | 1:1 | 1:1, 1:2, 1:3, 1:5, 1:10, 2:1, 3:1, 5:1, 10:1 (with F:M)
	3	N.E.S.	37, 75, 150, 300, 600 | 12, 25, 50, 100, 200, 400 | 500	3:2:1 (with M:F:M) | 1:1:1 (with M:F:M | 1:1:1, 3:1:1, 3:1:0.5, 3:1.5:1 (with F:M:F)
	4	N.E.S.	62, 125, 250, 500, 1000 | 12, 25, 50, 100, 200, 400 | 500	4:3:2:1 (with F:M:F:M) | 1:1:1:1 (with F:M:F:M) | 1:1:1:1, 3:1:1:1, 3:2:1:0.5 (with F:M:F:F)

Laboratory PGS (version) *STR Kit* ABI CE	NoC Range	# samples	Total DNA Quantity Range (pg)†	Mixture Ratios Range†
Idaho State Police Services Forensic Laboratory (Meridian, ID) STRmix (v2.8) *Fusion 5C* ABI 3500	1	N.E.S.	7.8 to 4000	N/A
	2	N.E.S.	1000 | N.E.S.	25:1, 15:1, 9:1, 5:1, 3:1, 2:1, 1:1 | 19:1, 10:1, 5:1, 3:1, 1:1
	3	N.E.S.	N.E.S.	10:5:1, 3:2:1, 1:1:1
	4	N.E.S.	N.E.S.	10:5:2:1, 4:3:2:1, 1:1:1:1
Jefferson County Regional Crime Laboratory (Golden, CO) STRmix (v2.6/2.6.3) *GlobalFiler* ABI 3500	1	25	16, 31, 63, 125, 250, 500, 2000, 4000, 8000	N/A
	2	10 | 31	262.5 to 400 | 131 to 2100	20:1, 10:1, 5:1, 3:1, 2:1, 1:1 | 20:1, 10:1, 5:1, 3:1, 1:1
	3	21 | 12*	37 to 1600 | N.E.S.	10:5:1, 3:2:1, 1:1:1 | 9:1:1, 5:1:1, 3:1:1, 1:1:1
	4	16 | 24*	50 to 1800 | N.E.S.	10:5:2:1, 4:3:2:1, 1:1:1:1 | 9:5:1:1, 9:3:1:1, 9:1:1:1, 5:1:1:1, 3:1:1:1, 1:1:1:1
	5	4 | 4*	700, 700, 702, 705 | N.E.S.	10:5:1:1:1, 5:4:3:2:1, 4:3:1:1:1, 1:1:1:1:1 | 9:5:5:1:1, 9:5:3:3:1, 9:3:3:1:1, 4:2:1:1:1
Las Vegas Metropolitan Police Department (Las Vegas, NV) STRmix (v2.6/2.6.3) *Investigator 24plex* 28 cycles ABI 3500xL	1	Not mentioned	Not mentioned	N/A
	2	204	50, 200, 500, 1000, 1500, 2000	8:1, 6:1, 5:1, 4:1, 3:1, 2:1, 1:1
	3	64	50, 200, 500, 1000, 1500	15:2:1, 12:2:1, 10:5:1, 9:2:1, 6:2:1, 3:2:1, 1:1:1
	4	65	50, 200, 500, 1000, 1500	15:1:1:1, 12:1:1:1, 10:5:2:1, 9:1:1:1, 6:1:1:1, 4:3:2:1, 1:1:1:1
	5	15	50, 200, 500, 1000, 1500	16:1:1:1:1, 12:1:1:1:1, 8:1:1:1:1
Los Angeles County Sheriff's Department (Los Angeles, CA) STRmix (v2.5.11) *Fusion 6C* 29 cycles ABI 3500	1	22	13, 25, 27, 36, 48, 50, 60, 100, 114, 250, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 6840	N/A
	2	32	200, 1000	80:1, 60:1, 40:1, 20:1, 10:1, 5:1, 4:1, 3:1, 2:1, 1:1
	3	34	75, 150, 300, 600, 900, 1000	20:10:1, 20:5:1, 20:1:1, 10:1:1, 5:1:1, 1:1:1
	4	34	100, 200, 300, 400, 800, 1000, 1200	10:5:5:1, 10:5:1:1, 10:1:1:1, 5:5:5:1, 4:3:2:1, 1:1:1:1

Laboratory PGS (version) *STR Kit* ABI CE	NoC Range	# samples	Total DNA Quantity Range (pg)[†]	Mixture Ratios Range[†]
Michigan State Police Forensic Science Division (Lansing, MI) STRmix (v2.3.07) *PowerPlex Fusion* 30 cycles ABI 3500	1	6 (+31)	25, 50, 75, 150, 500, 600	N/A
	2	18 (+12)	500, 1000, 3000	10:1, 7.5:1, 5:1, 2.5:1, 1:1, 4:1
	3	22 (+32)	500, 1000	20:5:1, 20:1:1, 10:10:1, 10:7.5:1, 10:5:1, 10:2.5:1, 10:2:1, 10:1:1, 5:1:1, 4:1:1, 5:5:1, 4:4:3, 4:4:1, 3:2:1, 2:2:1, 1:1:1
	4	19 (+10)	260, 500, 580, 780, 1000, 1170	20:5:1:1, 20:1:1:1, 10:10:5:1, 10:10:2:1, 10:10:1:1, 10:5:1:1, 10:2:1:1, 10:1:1:1, 5:5:5:1, 4:3:2:1, 2:2:2:1, 1:1:1:1
NYC OCME Forensic Biology Laboratory (New York City, NY) STRmix (v2.4) *PowerPlex Fusion* 29 cycles ABI 3130	1	3 | 30 | 5	10, 25, 50, 100, 200 | 750, 1000, 1500 | 2000	N/A
	2	N.E.S.	500	15:1, 10:1, 4:1, 2:1, 1:1
	3	N.E.S.	N.E.S.	N.E.S.
	4	N.E.S.	N.E.S.	N.E.S.
NYC OCME Forensic Biology Laboratory (New York City, NY) STRmix (v2.7) *Fusion 5C* 29 cycles ABI 3500xL	1	N.E.S.	3.25, 7.5, 15, 25, 37.5, 50, 75, 100, 125, 250, 525, 750, 1000, 2000	N/A
	2	N.E.S.	37.5 to 750	N.E.S.
	3	N.E.S.	37.5 to 750	N.E.S.
	4	N.E.S.	37.5 to 750	N.E.S.
	5	N.E.S.	37.5 to 750	N.E.S.

Laboratory PGS (version) *STR Kit* ABI CE	NoC Range	# samples	Total DNA Quantity Range (pg)[†]	Mixture Ratios Range[†]
Oregon State Police Forensic Services Division (Portland, OR) STRmix (v2.3, v2.4) *GlobalFiler* 29 cycles ABI 3500xL	1	N.E.S.	10, 20, 30, 40, 50, 60, 70, 90, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500	N/A
	2	N.E.S.	100, 500, 1000	70:1, 60:1, 50:1, 40:1, 30:1, 20:1 Siblings (3:1, 2:1, 1:1)
	3	N.E.S.	100, 500, 1000	70:20:10, 60:20:20, 50:25,25, 45:45:10, 80:10:10 Parent/Offspring (70:20:10, 20:60:20, 45:45:10, 10:80:10, 25:50:25)
	4	N.E.S.	100, 500, 1000	40:40:10:10, 50:25:12.5:12.5, 60:20:10:10, 70:20:5:5 2 Parents/2 Offspring (40:40:10:10, 12.5, 12.5, 50, 25, 10:60:20:10, 70:5:20:5)
Palm Beach County Sheriff's Office (W. Palm Beach, FL) STRmix (v2.4.06) *PowerPlex Fusion 5C* - 30 cycles ABI 3500xl	1	N.E.S.	30, 60, 125, 250, 500	N/A
	2	N.E.S.	100, 250, 500 | 100, 250, 500, 1000	19:1, 10:1, 5:1, 2.5:1, 1:2.5, 1:5, 1:10, 1:19 | 1:1
	3	N.E.S.	100, 250, 500, 1000 | 100, 250, 500, 1000	1:1:8, 6:3:1, 5:5:1, 1:3:3 | 1:1:1
	4	N.E.S.	100, 250, 500, 1000 | 100, 250, 500, 1000	4:4:1:1, 1:1:3:6, 1:3:3:9 | 1:1:1:1
Palm Beach County Sheriff's Office (W. Palm Beach, FL) STRmix (v2.6.2) *PowerPlex Fusion 6C* - 29 cycles ABI 3500xl	1	N.E.S.	12, 25, 50, 100, 200, 400	N/A
	2	N.E.S.	100, 250, 500, 1000	20:1, 10:1, 5:1, 2:1, 1:2, 1:5, 1:10, 1:20
	3	N.E.S.	100, 250, 500, 1000	10:5:1, 8:1:1, 3;2:1, 1:1:1
	4	N.E.S.	100, 250, 500, 1000	10:5:2:1, 9:3:3:1, 6:3:1:1, 4:4:1:1, 4:3:2:1, 1:1:1:1
Sacramento County District Attorney's Crime Laboratory (Sacramento, CA) STRmix (v2.4) *Fusion 6C* ABI 3500xL ("various amounts of allele sharing across the different loci")	1	10	10, 25, 50, 100, 200, 1000, 8000	N/A
	2	31 (53?)	500	20:1, 10:1, 5:1, 3:1, 2:1, 1:1
	3	27 (44?)	N.E.S.	10:5:1, 4:2:1, 4:1:1, 3:3:1, 3:2:1, 1:1:1
	4	23 (35?)	N.E.S.	10:5:2:1, 5:3:1:1, 5:1:1:1, 4:4:1:1, 4:3:2:1, 1:1:1:1
	5	6	N.E.S.	10:1:1:1:1, 15:1:1:1:1

Laboratory PGS (version) *STR Kit* ABI CE	NoC Range	# samples	Total DNA Quantity Range (pg)[†]	Mixture Ratios Range[†]
San Diego Police Department Crime Laboratory (San Diego, CA) STRmix (v2.3.06) *GlobalFiler* 29 cycles ABI 3500	1	N.E.S.	N.E.S.	N/A
	2	42	N.E.S.	8:1, 5:1, 2:1, 1:1, 1:2, 1:5, 1:8
	3	66	N.E.S.	33.3:33.3:33.3, 70:20:10, 60:30:10, 50:40:10, 50:30:20, 45:45:10, 40:40:20, 35:35:30, 60:20:20, 50:25:25, 40:30:30
	4	66	N.E.S.	25:25:25:25, 60:20:10:10, 50:20:20:10, 70:10:10:10, 40:20:20:20, 40:40:15:5, 35:35:20:10, 40:40:10:10, 35:35:25:5, 30:30:20:20, 30:30:30:10
	5	12	N.E.S.	20:20:20:20:20, 60:10:10:10:10
Virginia Department of Forensic Science (Richmond, VA) TrueAllele Casework (v3.25.4441.1) *PowerPlex 16* ABI 3130xl	1	17	10, 30	N/A
	2	18	N.E.S.	N.E.S. (mixture weight in Table 1)
	3	15	N.E.S.	N.E.S. (mixture weights in Table 2)
	4	7	N.E.S.	N.E.S. (mixture weights in Table 3)
Department of Forensic Sciences (Washington, DC) STRmix (v2.3) *Identifiler Plus* ABI 3130xl	1	N.E.S.	25, 50, 100 200, 400, 2000, 4000, 8000	N/A
	2	N.E.S.	500, 1000	20:1, 15:1, 10:1, 7:1, 3:1, 1:1
	3	N.E.S.	N.E.S.	20:10:1, 20:1:1, 10:1:1, 5:2.5:1, 3:1.5:1, 3:1:1
	4	N.E.S.	N.E.S.	10:10:10:1, 10:10:1:1, 10:5:2:1, 10:1:1:1, 5:5:5:1, 5:5:1:1, … 1:1:1:1
Department of Forensic Sciences (Washington, DC) STRmix (v2.4) *GlobalFiler* 29 cycles ABI 3500	1	32	6, 8, 12, 15, 23, 31, 47, 63, 94, 125, 188, 250, 375, 500, 750, 1000	N/A
	2	42	300, 600	25:1, 20:1, 15:1, 10:1, 7:1, 5:1, 3:1, 2:1, 1:1
	3	20	200, 500, 900	20:10:1, 10:5:1, 10:2:1, 5:1:1, 3:2:1, 3:1:1
	4	20	100, 200, 400, 600, 700, 800, 900, 1000	20:10:1:1, 20:5:2:1, … 2:2:2:1, 2:2:1:1
	5	20	300, 600, 1000	20:1:1:1:1, 10:10:10:10:1, … 5:1:1:1:1, 3:1:1:1:1, 2:2:2:1:1

Laboratory PGS (version) *STR Kit* ABI CE	NoC Range	# samples	Total DNA Quantity Range (pg)†	Mixture Ratios Range†
Wisconsin State Crime Laboratory (Milwaukee & Madison, WI) STRmix (v2.6) *Fusion 6C* 28 cycles ABI 3500xL	1	56	10, 25, 50, 100, 200, 4500, 5000, 5500	N/A
	2	N.E.S.	100, 200, 375, 750, 1000, 1500	40:1, 20:1, 10:1, 5:1, 3:1, 2:1, 1:1
	3	N.E.S.	100, 200, 375, 750, 1000, 1500	12:2:2, 10:5:1, 7:7:2, 4:2:1, 3:2:1, 2:2:1, 1:1:1

5.1. **Comments on Publicly Accessible PGS Internal Validation Summaries**

For the 20 publicly accessible PGS internal validation summaries described in Table S2.4, almost all (95%; 19 of 20) of the studies involved various versions of STRmix and different STR typing kits and one study assessed TrueAllele and PowerPlex 16. *These documents provide information from earlier versions of PGS systems that may not be currently used in the associated laboratories.* Half of these documents (10 of the 20 examined) were validation summaries performed and described by the Institute of Environmental Science and Research Ltd (ESR), the commercial provider of STRmix.

The validation summaries describe results from single-source samples as well as two-person and three-person mixtures and sometimes four- or five-person mixtures with contributor ratios ranging up to 25 times the quantity of the smallest contributor for two-person mixtures and up to 10 times the quantity of the smallest contributor for three-person mixtures.

There have been thousands of samples studied based on what is cited in this supplemental document to the main NIST report. However, the data reported in these studies is (1) generally not accessible to others outside of the groups generating it and (2) is potentially covering only a small portion of case scenarios where PGS is being applied in forensic casework.

5.1.1. **Additional STRmix Data**

Several of the public commenters on our June 2021 draft report provided their laboratory validation summaries or further information. For example, the Michigan State Police provided additional details on their internal validation studies (PC38 in 8351-draft PCs), the Las Vegas Metropolitan Police Department provided copies of their Investigator 24plex STRmix v2.6 validation summary prepared by ESR (PC12 in 8351-draft PCs; see also LVMPD 2020), and the Idaho State Police provided their PowerPlex Fusion 5C STRmix v2.8 validation summary prepared by ESR (PC63 in 8351-draft PCs).

In response to the draft report (see PC10 and PC53 in NISTIR 8351-draft PCs), ESR provided STRmix data on their website (ESR 2021). The provided information is in zip files related to six publications (Bille et al. 2019, Bright et al. 2018, Buckleton et al. 2018, Buckleton et al. 2020, Kelly et al. 2018, Lin et al. 2020). For example, data associated with the 2020 article on mixed

DNA profiles from a mother, father, and child trio (Lin et al. 2020) contains a ReadMe file that states: "Due to privacy reasons, the genotypes of the known contributors and EPGs are not shared. The genotypes of the non-contributors are given. Template and degradation for each profile and allele sharing for known and non-contributors is also provided." One of the associated data files contains 10,000 simulated profiles with 23 STR markers found in GlobalFiler and PowerPlex Fusion 6C kits. The other data file in this folder contains 800,507 rows of information with assigned LR values using the 10,000 simulated profiles for non-contributor testing under varying levels of DNA template degradation and allele sharing.

A spreadsheet file was provided related to the reanalysis of the NIST MIX13 case 5 using STRmix and EuroForMix (Buckleton et al. 2018) that contained three worksheets: (1) 10,000 simulated DNA profiles with alleles from the 15 STR markers in the Identifiler kit for non-contributor testing, (2) STRmix assigned LR values using the 10,000 simulated profiles, and (3) EuroForMix assigned LR values using the 10,000 simulated profiles. According to the accompanying ReadMe file, these were the data points used to construct Figure 2 in the original article (Buckleton et al. 2018). In this case, the original EPGs and reference profiles from NIST MIX13 are available[6] on the NIST STRBase website.

The ESR STRmix data release provides an example of the challenges in understanding what data can and should be shared to enable independent analysis. The primary information provided by ESR involved simulated genotypes and results for non-contributor testing performed. This information has limited value without the true contributor genotypes, which were not released in some cases due to privacy concerns. Publicly accessible data need to be published with metadata to enable use by others. The structure of the provided data files, which may make perfect sense to the provider, may not be clear to potential users who wish to conduct an independent analysis of results described in a published article or an internal validation study. More research and collaboration will be needed (see Key Takeaway #4.7 in NISTIR 8351) to arrive at standardized formats for sharing, comparing, and analyzing the detailed data that can be generated by PGS systems.

5.1.2. **Some Observations and Areas for Improvement**

When something is designated as N.E.S. (not explicitly stated) in the referenced publicly accessible source, it points to a potential need for the future development of standardization of data reporting. For example, many internal validation studies described in Table S2.4 do not clearly state the number of samples tested, making it difficult to assess the extent of the studies. *Moreover, given that the details were NIST-extracted rather than provided in the format listed, some information may have been missed or misunderstood.* NIST did not reach out specifically to the laboratories mentioned below and made any observations based solely on publicly accessible information. The availability of this public data is greatly appreciated, and any examples described are to encourage more detail and context in future validation summaries.

[6] See https://strbase.nist.gov/NIST_Resources/Interlabortory_Studies/Mix_13.zip (accessed October 21, 2024)

Sometimes the provided sample numbers do not agree in different parts of the validation summary. For example, in the Sacramento County District Attorney's Crime Laboratory Internal Validation of STRmix v2.4 report dated March 8, 2017, Table 3 (p. 10) in Section D of the ESR-derived report describes the number of samples examined with specificity and sensitivity tests as 31 two-person, 27 three-person, 23 four-person, and 6 five-person mixtures. However, Section S of the same report titled "Mixture summary" includes 53 two-person mixtures in Tables 13-15 (pp. 30-33), 44 three-person mixtures in Tables 16-18 (pp. 34-37), and 35 four-person mixtures in Tables 19-21 (pp. 38-41). For this validation summary entry in Table S2.4, the different values are both included with the larger number in parentheses with a question mark. Also, in Table S2.4, updated information for the Michigan State Police entry based on public comments received (see PC38 in NISTIR 8351-draft PCs) were included as "(+sample number)" in the sample number column along with additions to the total DNA quantity and mixture ratios where appropriate.

Each laboratory appears to have explored slightly different factors in their validation studies. Ten of these studies examined four-person mixtures involving contributor ratios spanning 17:1:1:1 to 10:10:5:1 to 4:3:2:1 to 1:1:1:1. Many studies were conducted with total DNA quantities in the range of 500 pg to 1000 pg although minor contributor quantities were sometimes in the range of single-cell analysis (6 pg) where significant allele drop-out would be expected.

Seven of the 20 studies in Table S2.4 describe the examination of five-person mixtures, including 12 samples reported by the San Diego Police Department Crime Laboratory and 20 samples reported by the Washington DC Department of Forensic Sciences. Information on DNA quantities examined, mixture ratios studied, and degree of allele sharing in these five-person mixture samples was not always explicitly stated in the referenced public sources. Additional data exploring five-person mixtures (and other mixtures examined) may exist within these and other forensic laboratories; however, as described in the main report (NISTIR 8351), the summaries here considered only publicly accessible data[7].

Sometimes helpful information is missing in these validation summaries, such as:

- Number of samples examined of various types (e.g., 2p, 3p, or 4p mixtures)
- Assigned LR values with specific propositions used
- A description of the number of PCR cycles used
- Total amounts of DNA amplified (and per contributor)
- Degree of allele sharing among contributors (e.g., see Riman et al. 2024b)
- DNA profiles for the mixture contributors

[7] In their public comments, several forensic laboratories indicated their validation studies have been reviewed as part of their accreditation process (so they have had outside review just not continuous open access to the data), and they invited NIST team members to privately review their validation studies in a similar manner. Such an approach was not viewed as being a potential useful exercise in the context of our efforts. NIST evaluation of validation information would only move the needle from "trust the laboratories" to "trust the NIST authors who have looked at the laboratory data" rather than sharing publicly accessible data to facilitate independent review by any interested party.

As described in Chapter 1 of the main report, these are important factors that influence the complexity of DNA mixtures and/or the degree to which the reproducibility or accuracy of the results have been examined. Summaries typically do not provide data points (e.g., assigned LR values) and associated information and metadata (see Box 4.1 in NISTIR 8351) that could help assess the degree of reliability and performance under potential similar case scenarios.

Multiple public comments to our June 2021 draft report noted (1) that validation data consists of DNA profiles that are ethically confidential when informed consent for their use does not explicitly permit sharing them, and (2) any request for making validation data publicly accessible is a new requirement over and above mechanisms by which DNA validation studies are currently evaluated such as audits, court discovery, and Freedom of Information Act requests.

Maintaining legacy information from previous validation studies can also be important. The Idaho State Police (ISP) Forensic Services Laboratory has a website that shares their validation summaries[8]. However, the STRmix v2.8 Fusion 5C 3500 validation summary from the ISP Laboratory provided as part of the public comment period (see PC63 in NISTIR 8351-draft PCs) has been replaced by STRmix v2.9 PowerPlex Fusion 6C 3500 on their public website. Other groups maintain previous versions of validation studies (e.g., see websites for DC DFS and NYC OCME in Table S2.3).

5.1.3. **Studies Involving a High-Degree of Allele Sharing**

Biological relatives are expected to share more alleles with each other than with unrelated people (see Principle 6 in Chapter 2 of NISTIR 8351). If close biological relatives are expected to be present in DNA mixtures from casework samples, then it would be appropriate to design mixtures with these sample types and perform validation studies that explore the impact of allele sharing on assigned LR values. For this reason, examining the "sharing of alleles among contributors" is a requirement under the 2015 SWGDAM Guidelines for Validation of Probability Genotyping Systems (SWGDAM 2015, Guideline 4.1.6.5). Several publicly accessible PGS internal validation studies[9] from Table S2.3 have examined samples from close biological relatives with a high-degree of allele sharing, such as parent-child and sibling-sibling.

There are at least two issues to consider with potential DNA mixtures involving a high-degree of allele sharing from biological relatives. First, the complexity of the mixture itself and the ability to reliably deconvolute or separate mixture components. Second, the potential for a close biological relative who is a non-contributor to be falsely included (i.e., produce a large assigned LR value when this individual's DNA is *not* in the mixture). These two issues are somewhat interrelated as the ability to reliably deconvolute a mixture into various components – and thus assist with differentiating true contributors from non-contributors – depends in large measure on the complexity of the mixture itself. Allele drop-out and mixture ratios near 1:1 reduce the

[8] See https://isp.idaho.gov/forensics/validation-summaries/ (accessed October 21, 2024)
[9] Many of these PGS internal validation studies became publicly accessible in July 2021 when shared by the Brooklyn Defender Services on their website in response to public comment on the June 2021 draft NIST report (see PC36 in 8351-draft PCs).

ability to resolve contributors (i.e., more potential genotypes have to be considered at each STR locus, which lowers the overall assigned LR).

One laboratory constructed mixtures originating from two closely-related individuals as a "worst case scenario" as well as a four-person mixture containing "two related individuals at low levels" that was "further complicated by close relatives of the true donors being the non-donors being compared" (LVMPD, p. 7). When a mixture contained a total DNA quantity that was low (~250 pg) or possessed mixture ratios near 1:1, and were thus difficult to fully resolve, then "a close relative of the true contributor could be included purely by the spread of genotype combinations and the fact they are likely to possess some of the same alleles as the true donor(s)" (LVMPD, p. 7). The authors of this internal validation summary state: "It is recommended that caution be exercised if the case information suggests that relatives may be a consideration and where the profiles recovered are of low template/peak height. Nevertheless, STRmix was able to reliably exclude related non-contributors when interpreting profiles with high APH [average peak height] and greater distinction in mixture ratio" (LVMPD, p. 8).

This validation summary described the outcome of studies with related individuals as follows: "These samples drive a few H_d true LRs [i.e., LR values assigned to true non-contributors] above 1, ranging in magnitude from 10^3 to 10^{13}" (LVMPD, p. 7). Unfortunately, details were not provided in the publicly available validation summary of this assigned LR for a close relative non-contributor in the range of 10^{13} (10 trillion). Was this result from a mixture with a high-degree of allele sharing, with some contributors that were similar in mixture ratio, with allele drop-out of the true POI contributor, or all of these possibilities? This same summary also mentioned that the assigned LR for a known non-contributor that was unrelated was "approximately 10^3" and noted "As previous studies have demonstrated, adventitious matches may occur where, by chance, an individual possesses alleles that are represented in the mixture, and may be exacerbated if dropout has been proposed" (LVMPD, p. 7).

Other publicly accessible validation summaries contain similar results with studying close biological relatives and examining assignment of LR value(s) for related non-contributors. For example (see Figure 8 in Jefferson County, p. 24), a plot of log(LR) versus average peak height with apparent three-person mixtures containing relatives showed assigned LRs for non-contributors that appear to be up to around 15 (i.e., an assigned LR of 10^{15}, one quadrillion) and true-contributors in this same set of mixtures as low as approximately -17 (i.e., an assigned LR of 10^{-17}). The discussion of this information in the validation summary included the following statements: "There were a number of non-contributors that gave LRs supporting H_p [i.e., LR>1]. LRs greater than 1,000 (log(LR)=3) were investigated further for the apparent two, three, four and five-contributor mixtures. In most examples the non-contributor was found to be a close relative (mother, father, sibling) of at least one of the known contributors. These individuals would be expected to have very similar DNA profiles and the correspondence to components of the mixture is not unexpected" (Jefferson County, p. 22). It would be helpful to have more details about these samples as this information could potentially help a laboratory develop appropriate limits to their PGS protocol depending on case circumstances involving relatives.

For mixtures consisting of contributors that are close biological relatives with a high-degree of allele sharing, the mixture may appear to have fewer contributors based on the maximum allele count method that is commonly used for NoC determination (see Table 9 in Brinkac et al. 2023). When the NoC determination is lower than the correct NoC, true contributors may be falsely excluded during the PGS analysis (e.g., Benschop et al. 2015b, Kelly et al. 2022a).

To what degree will DNA mixtures with contributors containing many shared alleles result in assignment of false positive, inclusionary LR values (i.e., LR>1) with DNA profiles from non-contributing close biological relatives? Based on what can be ascertained from publicly accessible validation studies and published research articles (e.g., Benschop et al. 2019a, Kalafut et al. 2022a, Kalafut et al. 2022b, Kelly et al. 2022b), the answer is that it depends on the mixture and the case circumstances.

5.1.4. **Examples of Claims and Recommendations Made**

Validation summaries may contain claims or share conclusions about observations made from the data examined. A sampling of some observations/claims extracted from various PGS internal validation summaries include:

- "As anticipated, as the input information (such as peak heights) decreases, the LR tends towards 1 (inconclusive) for both known contributors and known non-contributors" (LVMPD, p. 28).
- "At low template or high contributor number STRmix correctly and reliably reported that the analysis of the sample tends towards an uninformative or inconclusive LR" (Sacramento County, p. 10).
- "As expected, the likelihood ratio (LR) increased as the number of alleles detected increased in single source samples" (DFS Zoom Study, p. 3).
- "As expected, and demonstrated in the previous validation, higher-level contributors produce fully concordant inclusions/exclusions" (DFS Zoom Study, see p. 5 for 2-person mixtures, p. 8 for 3-person mixtures, p. 12 for 4-person mixtures, p. 14 for 5-person mixtures).
- "As with previous validations, caution is required when relatives are a consideration within certain mixture types, particularly when the mixtures are at a low level and the contributors are in roughly equal proportions" (LVMPD, p. 28).

One study described some general guidelines and limitations about assigned LR expectations from a PGS system based on the information collected using low quantities of DNA from GlobalFiler on the ABI 3500 with STRmix v2.4 (see DFS Zoom Study, p. 17):

- "Single source samples with results at less than 5 autosomal loci and average peak heights of less than 125 RFU are not expected to provide an LR above 100."
- "2-person mixtures where the lower-level contributor displays less than 10 unique minor autosomal alleles and average peak heights of less than 125 RFU are not expected to provide an LR above 100."
- "For 3- and 4-person mixtures, contributors with DNA amounts less than 200 may produce false inclusions. For 5-person mixtures, contributors with DNA amounts less than 375 may produce false inclusions."

- "These recommendations have been made using results from single amplification data. Replicate amplifications are expected to reduce the levels at which false inclusions have been observed."

Another study described their recommendations for implementing Investigator 24plex on the ABI 3500 with STRmix v2.6.3 into casework (LVMPD, p. 29):

- "The software may be used in the deconvolution and calculation of LRs for two-, three-, and four-person mixtures. The use of this software is *not approved* for use with mixtures that exhibit signs of being from five contributors."
- "With the exception of full single-source profiles where all homozygous loci are above the qualitative 400 RFU stochastic threshold and all heterozygous loci are above the 225 RFU drop-in cap, DNA profiles will be deconvoluted in STRmix prior to the calculation of an LR."
- "In the event the number of contributors to a profile is ambiguous, profiles should be run in STRmix using the lowest number which can most reasonably explain the DNA profile."
- "STRmix will be used to individually assess reference standards to determine whether they are included, excluded, or inconclusive as a contributor to the evidence profile. Manual comparisons for exclusions may only be performed when two or more loci resolve to 100% weight following STRmix deconvolution."
- "The LVMPD will consider an LR with an exponent falling between 10^3 and 10^{-3} as uninformative and will be reported as such. If an individual is determined to be uninformative, this individual will not be considered in combination with other individuals in alternative propositions."
- "The 99% 1-sided lower HPD [highest posterior density] interval will be used to report inclusions to partial single source, two, three, or four-person mixtures."
- "At minimum, STRmix interpretation must be attempted on the following sample types:
    - "Single source DNA profiles: contain at least one allele above the dye-specific analytical threshold at 6 or more loci (not to include Amelogenin or the DYS391 locus)
    - "Conditioned mixture DNA profiles: contain at least one minor/foreign allele above the dye-specific analytical threshold at 6 or more loci (not to include Amelogenin or the DYS391 locus)
    - "Non-conditioned mixture DNA profiles: contain at least one allele above the dye-specific analytical threshold at 8 or more loci (not to include Amelogenin or the DYS391 locus)" (LVMPD, p. 29).

Note the use of laboratory-selected thresholds, such as at least one allele above the dye-specific analytical threshold at 8 or more STR loci for a non-conditioned mixture DNA profile (LVMPD, p. 29), or expectations like an assigned LR will typically be less than 100 when the minor contributor in a 2-person mixture has less than 10 unique alleles and average peak heights of less than 125 RFUs (DFS Zoom Study, p. 17). While it is not always clear from publicly accessible information what data were used to make these decisions, these thresholds and expectations become part of laboratory interpretation protocols, which can vary between laboratories (e.g., Brinkac et al. 2023) and thus influence suitability assessments (Hicklin et al. 2023a) and outcomes of interlaboratory studies (see Section 2.5).

A key claim that is supported by data in these internal validation studies is that less information (e.g., from a partial DNA profile) leads to lower assigned LR values. However, *general trends do not reflect the reliability of a specific LR assignment*.

Some laboratories have established an uninformative zone when assigned LR values fall in this range, such as 0.01 to 100 (Sacramento County, p. 10) or 0.001 to 1000 (LVMPD, p. 29). Due to the fact that internal validation studies have shown false positives (an assigned LR>1 when H_2 is true) and false negatives (an assigned LR<1 when H_1 is true) in these LR ranges, this uninformative zone approach has been taken to reduce potential issues with false positives and false negatives. PGS systems may also generate run diagnostics to assist users is assessing how well the examined data fit to statistical models used in the algorithm (e.g., Russell et al. 2019).

In the end, a laboratory decision-maker decides whether sufficient data are available from an internal validation study to support results coming from a laboratory protocol. Some of the publicly accessible PGS validation summaries conclude with a statement from the DNA Technical Leader, such as

> The findings demonstrate that [a specific PGS system] is *suitable for its intended use* for the interpretation of DNA profiles generated from crime scene samples within [a specific laboratory system name].

The decision-maker, in this case the DNA Technical Leader, has determined that there is sufficient information to consider the PGS system ready for use. The confidence that others should have in this statement may not be clear without publicly accessible data for independent examination and confirmation.

6. Publicly Accessible Proficiency Test Results

Proficiency tests (PT) provide a means to assess participant performance and to examine trends in DNA interpretation methods. If PTs are representative of commonly seen casework in a forensic laboratory, then these results can also help assess what PCAST termed "validity as applied" (PCAST 2016). Due to the fact though that most PTs "are not blind (i.e., analysts know they are being tested), not well controlled (e.g., participation is voluntary, and analysts may or may not receive assistance from others), and not particularly realistic (e.g., samples are often pristine and recycled from previous tests)" (Koehler 2013), current proficiency tests in the words of one researcher "are not well-designed to assess error rates in realistic settings" (Koehler 2013). However, information from even imperfect PTs may be helpful in understanding variation across participants.

As described in Chapter 3 of the main report (NISTIR 8351), Collaborative Testing Services, Inc. (CTS) is currently the only proficiency test provider to publicly share results. These results are coded to anonymize participants and yet permit a view of variation across individual submissions. In each of the CTS PTs, four samples are provided (either as samples or profiles): Item 1 and Item 2 serve as references for comparison to "evidence" Item 3 and Item 4. CTS also provides a mock case scenario for context. Participants conduct their analyses and interpretations according to their laboratory protocols and report their results.

For each item, participants return results for (1) body fluid screening (e.g., "positive," "negative," "inconclusive," or "not tested" for the presence of blood along with listing test(s) conducted), (2) allele calls for autosomal STR loci analyzed with one or more STR kits (and Y-chromosome STR loci and mitochondrial DNA sequencing, if performed), (3) interpretation, and (4) additional comments that may assist in review of their results. A differential extraction (see Box S1.1 in NISTIR 8351sup1) can be performed to separate DNA components into sperm and epithelial fractions.

Interpretation typically involves answering a question like: "Based on results obtained from DNA analysis, could the Victim (Item 1) and/or the Suspect (Item 2) be a contributor to the questioned samples (Item 3 and Item 4)?" Thus, the assessment is simply "Yes" or "No" (i.e., inclusion or exclusion) and does not include a statistical evaluation of the strength of evidence. Some participants may respond with "inconclusive" or "no interpretation" as well. The summary report from CTS provides manufacturer information about how the samples were created along with the "correct" result, which is determined by consensus of participants. A minimum of 10 participants is required for a result (e.g., genotype at an STR locus) to be considered a consensus value.

Information from CTS publicly available proficiency test results has been extracted and summarized as part of this review. These PT results are categorized into three sections below depending on their sample source or requested output: (1) from biological samples (see Table S2.5), (2) from electropherograms (EPGs) (see Table S2.6), and (3) where a probabilistic genotyping software (PGS) output with LR values was reported (see Tables S2.7 to S2.11).

With some PT datasets, a subset of the participants reported whether a PGS system was used to assist in their DNA mixture interpretation. This information was included as a column in

Table 4.6 of our initial draft report released in June 2021. Because it was unclear whether all participants provided their PGS use in an equivalent manner across the various proficiency tests, the percent participants using PGS column has been removed from Table S2.5 below. Removing this information helps avoid any confusion or potential misunderstanding that the CTS PT results may be linked to PGS reliability as inferred in a publication commenting on our draft report (Bille et al. 2022).

6.1. **CTS Proficiency Test Results from Biological Samples**

Since 2013, CTS has provided proficiency test results involving single-source blood samples and mixtures of blood and blood or blood and semen. Thus, these PTs involving biological samples may or may not contain DNA mixtures. Table S2.5 provides a NIST-extracted summary of 109 datasets from biological samples that were publicly accessible before this review was finalized.

The CTS Test Number column contains three different types of PTs, which are indicated in parentheses. First, DNA Mixture PTs, which are offered twice a year, are numbered using the last two digits of the year and the test number, which is 581 or 586 from 2013 through 2016 (e.g., 16-581) and 5801 or 5806 starting in 2017 (e.g., 17-5801). Second, DNA Semen PTs, which are offered twice a year, are also numbered by the year and the test number, which is 582 or 584 from 2014 through 2016 and 5802 or 5804 starting in 2017. Third, Forensic Biology PTs, which are offered six times per year, are numbered by the year and the test number, which is 571 to 576 from 2014 through 2016 and 5701 to 5706 starting in 2017. More participants typically utilize the Forensic Biology PT compared to the other two categories.

The remaining four columns in Table S2.5 contain the number of participants in the particular PT, a summary of the sample source for Items 3 and 4 (which are the mock evidence samples), and any incorrect results that were reported. Mock evidence samples provided by CTS (Item 3 or Item 4) include single-source blood (B) samples and blood/blood (B/B) or blood/semen (B/S) mixtures, which are typically mixed in equal parts (i.e., a 1:1 sample ratio). False exclusion or false negative (FN) results involve a participant reporting an exclusion of DNA results from a provided reference sample that was present in the evidence sample. False inclusion or false positive (FP) results involve a participant reporting an inclusion of DNA results from a reference sample that was not present in the evidence sample.

For these sets of CTS proficiency tests (Forensic Biology, DNA Mixture, and DNA Semen), success is judged on appropriate inclusion or exclusion of provided reference samples (Item 1 and Item 2). LR values are not required and typically are not provided with most participant responses.

Table S2.5. NIST-extracted summary of 109 available data sets from Collaborative Testing Services (CTS) Forensic Biology, DNA Mixture, and DNA Semen proficiency tests between 2013 and 2024.

CTS Test Number	Number of Participants	Samples Provided (sample ratio noted)		Any Incorrect Results
		Item 3	*Item 4*	
13-581 (DNA Mixture)	129	B/S (1:1)	B	2 FN
13-586 (DNA Mixture)	118	B	B/S (1:1)	--
14-571 (Forensic Biology)	778	B	B	5 FN, 3 FP
14-572 (Forensic Biology)	603	B	B/S (1:1)	1 FN
14-573 (Forensic Biology)	357	B/B (1:1)	B	3 FP
14-574 (Forensic Biology)	756	B/S (1:1)	B	1 FN
14-575 (Forensic Biology)	611	B/S (1:1)	B	1 FN
14-576 (Forensic Biology)	327	B	B/S (1:1)	3 FN
14-582 (DNA Semen)	171	B	B/S (1:1)	--
14-584 (DNA Semen)	169	B	B/S (1:1)	5 FN
14-581 (DNA Mixture)	131	B	B/S (1:1)	--
14-586 (DNA Mixture)	142	B/S (1:1)	B	4 FN
15-571 (Forensic Biology)	727	B/S (1:1)	B	--
15-572 (Forensic Biology)	631	B/B (1:1)	B	--
15-573 (Forensic Biology)	351	B	B/S (1:1)	1 FN, 1 FP
15-574 (Forensic Biology)	675	B	B	--
15-575 (Forensic Biology)	611	B	B/S (1:1)	1 FN
15-576 (Forensic Biology)	320	B	B	--
15-582 (DNA Semen)	179	B/S (1:1)	B	1 FN
15-584 (DNA Semen)	160	B	B/S (1:1)	--
15-581 (DNA Mixture)	145	B/S (1:1)	B	3 FN
15-586 (DNA Mixture)	121	B/S (1:1)	B	--
16-571 (Forensic Biology)	697	B	B	1 FN, 1 FP
16-572 (Forensic Biology)	659	B/S (1:1)	B	3 FN, 1 FP
16-573 (Forensic Biology)	360	B	B	--
16-574 (Forensic Biology)	615	B	B/S	1 FN
16-575 (Forensic Biology)	632	B/B (1:1)	B	--
16-576 (Forensic Biology)	329	B/S (1:1)	B	1 FP
16-582 (DNA Semen)	174	B	B/S (1:1)	1 FN
16-584 (DNA Semen)	188	B	B/S (1:1)	3 FN
16-581 (DNA Mixture)	142	B	B/S (1:1)	2 FN
16-586 (DNA Mixture)	144	B/B (1:1)	B/S (1:1)	3 FN
17-5701 (Forensic Biology)	672	B	B	--
17-5702 (Forensic Biology)	660	B	B	--
17-5703 (Forensic Biology)	348	B	B/S (1:1)	3 FN
17-5704 (Forensic Biology)	671	B/S (1:1)	B	--
17-5705 (Forensic Biology)	594	B/S (1:1)	B	1 FN, 2 FP
17-5706 (Forensic Biology)	327	B/B (1:1)	B/B (1:1)	1 FN, 1 FP
17-5802 (DNA Semen)	187	B	B/S (1:1)	--
17-5804 (DNA Semen)	194	B/S (1:1)	B	1 FN
17-5801 (DNA Mixture)	179	B/S (1:1)	B/S (1:1)	1 FN
17-5806 (DNA Mixture)	167	B/S (1:1)	B/B (1:1)	--
18-5701 (Forensic Biology)	683	B/B (1:1)	B	1 FN, 1 FP
18-5702 (Forensic Biology)	651	B	B/S (1:1)	1 FN
18-5703 (Forensic Biology)	359	B	B/S (1:1)	--

CTS Test Number	Number of Participants	Samples Provided (sample ratio noted)		Any Incorrect Results
		Item 3	*Item 4*	
18-5704 (Forensic Biology)	672	B/S (1:1)	B	1 FN
18-5705 (Forensic Biology)	624	B	B	--
18-5706 (Forensic Biology)	343	B/B (1:1)	B	--
18-5802 (DNA Semen)	226	B	B/S (1:1)	--
18-5804 (DNA Semen)	181	B/S (1:1)	B	1 FN
18-5801 (DNA Mixture)	156	B	B/S (1:1)	3 FN, 1 FP
18-5806 (DNA Mixture)	178	B/S (1:1)	B/B (1:1)	--
19-5701 (Forensic Biology)	689	B	B/S (1:1)	--
19-5702 (Forensic Biology)	680	B	B/B (1:1)	--
19-5703 (Forensic Biology)	348	B	B	--
19-5704 (Forensic Biology)	696	B	B	1 FN, 1 FP
19-5705 (Forensic Biology)	705	B/S (1:1)	B	13 FN, 1 FP
19-5706 (Forensic Biology)	333	B/B (1:1)	B/S (1:1)	--
19-5802 (DNA Semen)	223	B	B/S (1:1)	--
19-5804 (DNA Semen)	166	B/S (1:1)	B	3 FN
19-5801 (DNA Mixture)	169	B/S (1:1)	B/B (1:1)	--
19-5806 (DNA Mixture)	171	B	B/S (1:1)	--
20-5801 (DNA Mixture)	235	B/B (1:1)	B/S (1:1)	--
20-5806 (DNA Mixture)	202	B	B/S (1:1)	--
20-5701 (Forensic Biology)	671	B	B	--
20-5702 (Forensic Biology)	673	B/S (1:1)	B	6 FN
20-5703 (Forensic Biology)	345	B/B (1:1)	B	--
20-5704 (Forensic Biology)	706	B	B/S (1:1)	1 FN, 1 FP
20-5705 (Forensic Biology)	642	B	B	--
20-5706 (Forensic Biology)	316	B	B/S (1:1)	--
20-5802 (DNA Semen)	207	B/S (1:1)	B	--
20-5804 (DNA Semen)	186	B/S (1:1)	B	--
21-5701 (Forensic Biology)	622	B	B/B (1:1)	--
21-5702 (Forensic Biology)	736	B	B/S (1:1)	1 FN
21-5703 (Forensic Biology)	357	B	B/S (1:1)	1 FN
21-5704 (Forensic Biology)	575	B/S (1:1)	B	3 FN
21-5705 (Forensic Biology)	654	B	B	--
21-5706 (Forensic Biology)	334	B/B (1:1)	B	--
21-5802 (DNA Semen)	209	B	B/S (1:1)	1 FN, 1 FP
21-5804 (DNA Semen)	210	B	B/S (1:1)	5 FN
21-5801 (DNA Mixture)	167	B/S (1:1)	B	--
21-5806 (DNA Mixture)	222	B/S (1:1)	B/B (1:1)	--
22-5701 (Forensic Biology)	598	B	B	--
22-5702 (Forensic Biology)	709	B/B (1:1)	B	--
22-5703 (Forensic Biology)	363	B/B (1:1)	B/S (1:1)	--
22-5704 (Forensic Biology)	615	B/S (1:1)	B	--
22-5705 (Forensic Biology)	676	B/S (1:1)	B	--
22-5706 (Forensic Biology)	315	B/B (1:1)	B	3 FN, 2 FP
22-5802 (DNA Semen)	195	B	B/S (1:1)	1 FN, 2 FP
22-5804 (DNA Semen)	223	B/S (1:1)	B	--
22-5801 (DNA Mixture)	211	B/S (1:1)	B/B (1:1)	--
22-5806 (DNA Mixture)	204	B	B/S (1:1)	2 FN

CTS Test Number	Number of Participants	Samples Provided (sample ratio noted) *Item 3*	*Item 4*	Any Incorrect Results
23-5701 (Forensic Biology)	669	B/B (1:1)	B	--
23-5702 (Forensic Biology)	802	B	B/S (1:1)	--
23-5703 (Forensic Biology)	396	B/B (1:1)	B	--
23-5704 (Forensic Biology)	588	B/S (1:1)	B	10 FN
23-5705 (Forensic Biology)	719	B/B (1:1)	B	--
23-5706 (Forensic Biology)	353	B/S (1:1)	B	--
23-5802 (DNA Semen)	177	B/S (1:1)	B	1 FN
23-5804 (DNA Semen)	218	B	B/S (1:1)	1 FN
23-5801 (DNA Mixture)	156	B/B (1:1)	B/S (1:1)	1 FN
23-5806 (DNA Mixture)	201	B/S (1:1)	B	--
24-5701 (Forensic Biology)	642	B	B/S (1:1)	1 FN, 1 FP
24-5702 (Forensic Biology)	742	B	B/S (1:1)	5 FN, 18 FP
24-5703 (Forensic Biology)	378	B/B (1:1)	B/B (1:1)	--
24-5704 (Forensic Biology)	533	B	B/B (1:1)	--
24-5802 (DNA Semen)	202	B	B/S (1:1)	1 FN
24-5804 (DNA Semen)	250	B/S (1:1)	B	3 FN, 2 FP
24-5801 (DNA Mixture)	175	B/S (1:1)	B/B (1:1)	1 FN
TOTAL	**43,983**			**121 FN, 44 FP**

Across these 109 data sets, there were 175,932 possible responses (43,983 participants × two evidence items × two reference items). There were also "inconclusive" or "no response" decisions that are not always reflected in this data analysis. The ability to determine an exact denominator of a test is sometimes limited by how the data are tabulated and summarized by CTS.

From the publicly accessible information, there appears to be 121 false negatives and 44 false positives, which corresponds to 0.069% and 0.025%, respectively. It would be inappropriate though to use these error rates without further details, which are not always evident short of closer examination or seeking further details from the PT provider. For example, the CTS PT 24-5702 summary report in its Table 7 mentions 18 false positives[10]. The manufacturer shared in the summary report that consistent allelic results were seen for all STR loci, suggesting that allele drop-out from a sample with a low quantity of DNA was not likely a cause. A closer look found that 17 of the 18 false positives in this dataset were related to mitochondrial DNA results where the mtDNA haplotypes could not be distinguished. Thus, these findings do not relate to the ability to correctly perform DNA mixture interpretation. However, one of these false positives involved STR analysis by a participant using STRmix with GlobalFiler who incorrectly associated[11] the suspect (Item 2) with the evidence (Item 4), which was created using a 1:1

[10] See https://cts-forensics.com/reports/24-5702_Web.pdf (accessed October 15, 2024); these 17 participants, who incorrectly stated "Yes" to the Item3-to-Item1 comparison, all produced mtDNA results: 4AQTNX, 7HQ8FA, 863J6B, 8D8M3F, F79TCZ, F7J62B, FLHHPZ, KPM9T4, LD6E6U, LK8Z64, MCUYBT, MV3TQZ, MX7YAT, NGHX3U, P3VKBY, Y3LXQN.

[11] There may have been a transcriptional error by this participant (96J3GL) with their submission as it appears that the genotypes were correctly reported for the individual samples. The publicly accessible summary reports do not provide enough details to determine the root cause of this false positive.

mixture of blood and semen. Likewise, the 13 false negatives[12] seen in Table 7 of CTS PT 19-5705 appear to have reported mtDNA results also when investigated more closely. Unfortunately, the current CTS summary reports do not always provide the necessary granularity to make these details easily apparent.

6.2. CTS Proficiency Test Results from EPGs

Since 2013, CTS has provided DNA Interpretation proficiency test results involving analysis based only on EPGs provided. Participants in this PT therefore do not need to perform any measurements and laboratory work and can focus entirely on the interpretation part of the process (see NISTIR 8351, Figure 2.1). The EPGs come from a variety of STR typing kits. For example, in CTS Test 22-5882, EPG data were produced from seven kits (GlobalFiler, Investigator 24plex, PowerPlex Fusion 5C, PowerPlex Fusion 6C, Identifiler Plus, Yfiler Plus, and PowerPlex Y23. Table S2.6 provides a NIST-extracted summary of 22 available data sets in this category available on the CTS website before our report was finalized. Evidence profiles were designed from single individuals (single), two-contributor mixtures (2p), or three-contributor mixtures (3p) with the anticipated contributor ratios indicated in parentheses.

The number of participants for these PT datasets focusing on DNA interpretation from a provided EPG is much lower than the ones involving biological samples discussed in the previous section. Based on the publicly accessible CTS summary reports for each study, the number of false inclusions, false exclusions, inconclusives, and no responses are tabulated in Table S2.6. Some of participants provide additional details beyond their decision of inclusion or exclusion, such as interpretation guidelines used (e.g., analytical threshold, peak height ratio, stochastic threshold) and statistical analysis (e.g., LR values and population databases used). However, LR values were not requested by the PT provider for these types of tests and thus not commonly available to examine.

[12] See https://cts-forensics.com/reports/19-5705_Web.pdf (accessed October 15, 2024); these participants, who incorrectly stated “No” to the Item3-to-Item2 comparison, all produced mtDNA results: 2YFRBX, 347LK4, 7N7JWV, ALMXEV, EJHY8J, K22LFM, LVWDED, N68EYC, NG2UND, QEQQQ8, V4TKY4, WC39C2, YQXA39.

Table S2.6. NIST-extracted summary of 22 CTS DNA Interpretation proficiency tests between 2013 and 2024. If the contributor of interest (either Item 1 or Item 2) is present in the mixture, then this contribution to the mixture is underlined. Blue font indicates inclusion of a contributor in the evidence profile who is not a supplied reference profile ("Item 1" or "Item 2"). When responses were not all the same, they are subdivided (e.g., # false inclusions in the 15-588 row) and listed in the following order of comparison: Item3-to-Item1, Item3-to-Item2, Item4-to-Item1, Item4-to-Item2. Mixtures with more than two contributors (2p) are bolded (e.g., **3p**).

Year	CTS Test	Number of Participants	Item 3	Item 4	# False Inclusions	# False Exclusions	# Inconclusives	# No Response
2013	13-589	13	single	2p (4:1)	0	0	0	0
2014	14-588	20	2p (2:1)	single	0	0	0	0
2014	14-589	19	single	2p (2:1)	0	0	0	0
2015	15-588	19	single	2p (3:1)	0,1,0,0	0	0	0
2015	15-589	24	2p (1:4)	single	0	0	0	0
2016	16-588	20	2p (3:1)	2p (1:1)	0	0	1,3,0,3	0
2016	16-589	28	**3p** (2:1:2)	2p (4:1)	0	**1**,0,0,0	**2,4**,0,0	**1**,0,1,0
2017	17-588	21	**3p** (1:2:1)	2p (1:3)	0	0	**4,2**,1,0	**3**,0,3,0
2017	17-589	19	2p (1:4)	**3p** (5:1:3)	0	0,0,0,**1**	0,0,**2,4**	0
2018	18-588	25	2p (1:1)	2p (3:1)	0	0	0,0,3,0	0
2018	18-589	36	2p (3:1)	**3p** (6:3:1)	0	0	0,0,**12,12**	0
2019	19-588	28	**3p** (4:1:2)	2p (1:4)	0	0	**1,9**,0,0	0
2019	19-589	38	2p (2:3)	**3p** (5:2:2)	0	0	0,0,**7,9**	0
2020	20-5881	43	**3p** (5:1:3)	2p (4:1)	0	0	**7,9**,0,0	0
2020	20-5882	55	2p (5:2)	**3p** (5:2:1)	0	0	2,0,**17,16**	0
2021	21-5881	25	2p (1:3)	2p (3:5)	0,1,0,0	0	0	0
2021	21-5882	32	2p (1:3)	2p (2:1)	0,0,1,0	0	1,0,0,0	0,0,1,1
2022	22-5881	30	2p (2:1)	**3p** (5:3:2)	0	0	0,1,**2,1**	0
2022	22-5882	35	**3p** (1:3:1)	2p (1:3)	0	0,**13**,0,0	0,**11**,0,0	0
2023	23-5881	36	2p (1:1)	2p (2:1)	0	0	0	0
2023	23-5882	33	**3p** (2:1:1)	2p (1:2)	0	0	**4,3**,0,0	0
2024	24-5881	31	2p (1:1)	**3p** (1:1:2)	0	0	0,0,**3**,0	0
	TOTAL	**630**			**3**	**15**	**22,42,47,45 156**	**10**

These 22 sets of DNA Interpretation PT results provide 2520 responses (630 participants × two evidence items × two reference items). The responses include three (0.12%) false inclusions, 15 (0.60%) false exclusions, 156 (6.2%) inconclusive results, and 10 (0.40%) no responses.

The 3 false inclusions came from 2p mixtures where the reference profile of interest was not present in the provided evidence item. All 15 false exclusions came with 3p mixtures. With the 156 inconclusive results, 15 (9.6%) came from 2p mixtures and 141 (90.4%) came from 3p mixtures. For the 10 "no response" decisions, 6 (60%) came from 2p mixtures and 4 (40%) came from 3p mixtures.

Out of the 22 test sets examined, which involve 44 DNA "evidence" profiles, 11% (5) come from single-source samples, 61% (27) contain mixtures with two contributors ("2p"), and 27% (12)

involve three contributors ("3p"). Thus, there are some three-contributor mixtures unlike PT results described in the previous section (see Table S2.5).

6.3. **CTS Proficiency Test Results with Probabilistic Genotyping**

In 2022, CTS began offering a probabilistic genotyping proficiency test. These PTs provide an opportunity to examine variation across participants with the entire process from sample to reported result and LR value(s) obtained with a specific PGS system. Note that DNA extraction and PCR amplification efficiencies will impact levels of recovered DNA and thus potential LR values.

When this report was finalized in 2024, four sets of results from PGS PTs were available from 20 participants[13] with CTS Probabilistic Genotyping Test No. 22-5904/5, 48 participants[14] with CTS Probabilistic Genotyping Test No. 23-5901/2, 72 participants[15] with CTS Probabilistic Genotyping Test No. 23-5904/5, and 62 participants[16] with CTS Probabilistic Genotyping Testing No. 24-5901/2 (Table S2.7). With each test, two evidence samples (Item 3 and Item 4) that were artificially prepared mixture samples were compared to Item 1 (victim) and Item 2 (suspect). Possible responses for these two sets of comparisons were inclusion, exclusion, inconclusive, or no response. Thus, there were a total of 808 possible responses[17] (202 participants x 4 responses for each test).

Table S2.7 provides a NIST-extracted summary of the number of participants, complexity of the mixtures examined, and responses provided with these four CTS PGS proficiency tests. CTS tests use mixtures of blood and/or semen in liquid volume ratios with two (2p), three (3p), or four (4p) contributors. To form a 1:3 mixture, for example, one part of blood from one donor would be combined with three parts of blood from a second donor. However, different amounts of DNA-containing cells in each donor's blood means that the DNA mixture ratio may not match the prepared, reported, and desired mixture ratio. These liquid mixtures are aliquoted onto a sample substrate and dried before shipping. Tests 5901 and 5904 use a substrate of cotton swatches while Tests 5902 and 5905 use FTA micro card. In the 2024 test, there were 41 (66%) participants who used the 5901 cotton swatches while 21 (34%) used the 5902 FTA cards.

[13] See https://cts-forensics.com/reports/22-5904.5_Web.pdf (accessed October 15, 2024).
[14] See https://cts-forensics.com/reports/23-5901.2_Web.pdf (accessed October 15, 2024).
[15] See https://cts-forensics.com/reports/23-5904.5_Web.pdf (accessed October 15, 2024).
[16] See https://cts-forensics.com/reports/24-5901.2_Web.pdf (accessed October 15, 2024).
[17] These are not necessarily 202 unique practitioners as any participant could have taken more than one proficiency test over these set of tests. It is expected that groups of analysts from the same laboratory would report the same STR kit, PGS system, population database, and even the same way of expressing the propositions used and the LR value(s) obtained.

Table S2.7. NIST-extracted summary of CTS Probabilistic Genotyping proficiency tests between 2022 and 2024. When responses were not all the same, they are subdivided (e.g., row 1, # false exclusions) and listed in the following order of comparison: Item3-to-Item1, Item4-to-Item1, Item3-to-Item2, Item4-to-Item2. Mixtures with more than two contributors (2p) are bolded (e.g., **3p** and **4p**). If the contributor of interest (either Item 1 or Item 2) is present in the mixture, then this contribution to the mixture ratio is underlined.

CTS Test	Number of Participants	Item 3	Item 4	# False Inclusions	# False Exclusions	# Inconclusives	# No Response	Further Details
22-5904/5	20	**3p** (3:2:1)	2p (2:1)	0	0,1,0,0	0	0	Table S2.8
23-5901/2	48	2p (2:1)	**3p** (3:2:1)	0	0,2,0,0	0	0	Table S2.9
23-5904/5	72	2p (1:1)	2p (3:2)	0	0	0	0	Table S2.10
24-5901/2	62	**4p** (1:1:1:1)	**3p** (1:1:1)	0	0,0,7,0	1,0,3,0	5,0,5,0	Table S2.11
TOTAL	202	Four 2p mixtures Three **3p** mixtures One **4p** mixture		0/808 (<0.12%)	10/808 (1.2%)	4/808 (0.50%)	10/808 (1.2%)	

The first two of these probabilistic genotyping tests provided a three-person mixture with a 3:2:1 ratio and two-person mixture with a 2:1 ratio. Regarding whether the victim and/or the suspect could be a contributor to the questioned stains, potential response categories were inclusion, exclusion, inconclusive, no interpretation, and no response rather than an LR value with a specific set of propositions. On the surface, these results appear quite straightforward. As seen in Table S2.7, there were no false inclusions reported. Further details of the tests in the shaded cells, including individual participant LR values, may be found in Tables S2.8, S2.9, S2.10, and S2.11.

For the first PGS PT (see Table S2.8), there was one false exclusion (i.e., failure to include the victim, Item 1, in Item 4, which was a 2:1 mixture). However, based on accessible information[18] in the provided PT summary, it is unclear whether the participant exhibiting the alleged false negative even provided a response on whether the female victim's DNA was present on a stain from underwear the victim was wearing. Here the laboratory's policy could have impacted the approach taken by the PT participant.

For the second PGS PT (see Table S2.9), there were two false negatives (i.e., failure to include the victim, Item 1, in Item 4, which was a 3:2:1 mixture with the victim's DNA being the smallest contributor amount). An examination of the reported results[19] found that multiple alleles from

[18] In the additional comments section (Table 9 on p. 56), participant JZEB3C-5904 recorded: "Please note that the seminal portion only was profiled from item 4 as the aim was to ascertain the source of the semen detected given the case information provided."
[19] For example, in STR amplification kit(s) and results section (Table 3, p. 45), results for participant 2MZT8B-5901 were missing 17 alleles from the consensus profile (see p. 3). In the statistical analysis for item 4 section (Table 8, p. 81), this same participant reported: "…Victim visually excluded from comparable portion of the mixture (one of the contributors is a trace contributor)…." Thus, either not enough DNA from the mixture was amplified or use of a high analytical threshold resulted in the loss of a portion of the trace (victim) contributor.

the victim's DNA profile were missing in the Item 4 mixture profile. Here the participants' measurement and interpretation results impacted their interpretation capability.

For the third PGS PT (see Table S2.10), which involved only two-person mixtures, there were no false negatives or false positives reported. The assigned LR values reported spanned about 20 orders of magnitude from 10^9 to 10^{29}. This variation likely comes from some differences in STR kits utilized (e.g., Identifiler Plus with 15 STR markers versus GlobalFiler with 21 STR markers) and reporting caps (EWG 2024, pp. 85-87).

The fourth PGS PT (see Table S2.11) involved a four-person mixture where participants resulted seven false exclusions, four inconclusives, and 10 no responses. These responses along with the assigned LR values will be discussed later in this document.

In the first two columns of the NIST-extracted summaries for Tables S2.8 to S2.11, each participant's identifier (CTS WebCode) is tabulated along with their decision for the specific comparison shown at the top of the column (e.g., Item 3: Item 2). The STR kit utilized by each participant (column 3) is listed along with the PGS system, including the software version number when provided (column 4). The DNA mass concentration (ng/µL) (column 5) is reported as provided by the CTS summary, with concentration of the item obtained by quantitative PCR analysis multiplied by mixture proportion of that contributor, divided by 100. DNA proportions as percent values (column 6) are typically reported as computed by the probabilistic genotyping software[20]. Not all participants provided DNA mass concentration or DNA proportion values, nor is it clear whether the information from participants is being provided uniformly. The number of contributors (NoC) reported (column 7) and the population database used (column 9) were also tabulated. Finally, the assigned LR values (column 8) are listed with the same number of significant figures reported in the CTS summary table and have been adjusted to be on same scientific notation scale (e.g., 100 billion becomes 1E11).

[20] For example, see entry 9ME3CA-5904 in Table 9 (p. 114) of Test 23-5904/5 found at https://cts-forensics.com/reports/23-5904.5_Web.pdf (accessed October 15, 2024).

Table S2.8. NIST-extracted summary of responses provided in CTS Probabilistic Genotyping Test 22-5904/5 by 20 participants examining a 3-person mixture (based on mixing blood from three individuals in a 3:2:1 mixture). LR values are assigned from comparison of the suspect DNA profile (Item 2) to the mixture DNA profile (Item 3) and are normalized to a common scientific notation format (e.g., 1E11 rather than 10^{11} or 100 billion). Information that is missing from the publicly accessible source is listed as (--). *The listed GMID-X 1.6 is not a PGS system. LR = likelihood ratio; NoC = number of contributors; PGS = probabilistic genotyping software.

CTS WebCode	Item 3: Item 2	STR Kit	PGS	DNA mass conc (ng/µL)	DNA proportion (%)	NoC	LR Value for Item 2 in Item 3 (3:2:1)	Population Database
23ATZD	Included	PP21	STRmix v2.7	--	--	3	>1E11	Local
7LLZDK	Included	PP21	STRmix v2.8.0	--	--	3	>1E11	Local
8GAL9B	Included	ESI/ESX 17	EuroForMix	0.4469	71.10	3	3.5E6, 9.3E6, 1.2E9, 4.6E9	Local
C4A922	Included	PP21	STRmix v2.8.0	--	--	3	>1E11	Local
FNM3HX	Included	PP21	STRmix v2.8.0	--	--	3	>1E11	Local
H9ZM48	Included	PP21	STRmix v2.8	1.2372	69.00	3	1E11	Local
JBJ4WC	Included	GlobalFiler	STRmix	1.200	71.00	3	>1E9	Local
JMCKLD	Included	Invest24plex	STRmix v2.5.11	0.5820	71.00	3	1E15	FBI
JZEB3C	Included	NGM Select	STRmix	1.7614	50.00	3	at least 1E9	Local
KXDEH8	Included	GlobalFiler	GMID-X 1.6*	2.3900	66.50	3	1.92E11	Local
LGJ7YY	Included	Fusion 5C	TrueAllele	--	--	3	8E27, 6.7E28, 4E29	NIST
LZVVCW	Included	ESI Fast	EuroForMix	0.7400	61.56	3	1.45727856134815E20	Local
NH9AH2	Included	GlobalFiler	STRmix v2.6.3	1.5160	72.00	3	at least 1E11	--
RMPV4Z	Included	NGM Select	STRmix v2.7	--	--	3	1.0557E11	Local
TKUHUW	Included	PP21	STRmix v2.8.0	0.2720	68.00	3	1E11	Local
UXEKPU	Included	GlobalFiler	STRmix 2.5.11	--	--	3	1.92E28	Local
WVM3MY	Included	NGM	LRmix	6.6125	60.00	3	1.3E8	Local
X27TLZ	Included	ESI 17	STRmix	--	--	4	1E9	Local
X8X4LL	Included	ESI/ESX 17	EuroForMix	0.3029	69.20	3	3.2E10, 1.40E11, 8.4E10, 1.04E11	Local
ZXLNRJ	Included	Invest24plex	EuroForMix	0.3000	59.00	3	5.21E14, 2.99E27, 4.16E28, 2.49E42	Local

In this first PGS PT, 20 participants used 9 different STR kits: PowerPlex 21 (6x), GlobalFiler (4x), Investigator 24plex (2x), NGM Select (2x), ESI/ESX 17 (2x), ESI 17 (1x), ESI Fast (1x), NGM (1x), and PowerPlex Fusion 5C (1x). About two-thirds of participants (13 of 20; 65%) used STRmix followed by 20% (4 of 20) with EuroForMix, 5% (1 of 20) with TrueAllele, and 5% (1 of 20) with LRmix. One participant also reported using GMID-X 1.6, which is not a probabilistic genotyping system. Most of these participants (17 of 20; 85%) used a location identifying database, which is labeled "Local" in column 9. It is unclear what, if anything, can be made of the variation reported in DNA concentration although some participants may have a better DNA extraction procedure than others.

Of these 20 participants, 19 (95%) assumed the number of contributors to be 3 with the outlier (participant X27TLZ) assuming 4 individuals in the mixture. About half of the participants assigned LR values around 100 billion (1E11), which may well have been a cap on the reported results for their laboratory policies. The overall reported LRs ranged from 10^6 to 10^{42}, or 36 orders of magnitude. Multiple LR values were provided by some participants (e.g., LGJ7YY) that reflect LR assignments using various population groups.

The use of different propositions is likely responsible for much of this variation as can be seen in the last row in Table S2.8 (participant ZXLNRJ) where four LRs are provided and are dependent on conditioning with different combinations of possible suspect (S), victim (V), and unknowns (U) contributions: $\mathbf{10^{14}}$ (S+V+U/S+2U), $\mathbf{10^{27}}$ (S+2U/3U), $\mathbf{10^{28}}$ (S+V+U/V+2U), and $\mathbf{10^{42}}$ (S+V+U/3U). Some reported variation can also be seen with PCR replicates where participant 8GAL9B reported four LR values (3.5E6, 9.3E6, 1.2E9, 4.6E9) ranging from 35 million to 46 billion. Some also listed numerical values with qualifiers, such as greater than (>) or at least. Significant figures[21] shown are those provided in the CTS summary (e.g., LZVVGW reported 15 significant figures).

In the future, it would be helpful for the PT provider to request a greater level of detail and to use a standardized reporting format. While it is not satisfying to have missing information when extracting data from these CTS summary reports, the reality is that there is little to no standardization in reporting, nor is it currently required to report anything other than inclusion/exclusion with the comparisons made by these PT participants.

[21] A 2024 report recommends using at most one significant figure: "Recommendation 4.2: To avoid conveying an unsupported level of precision, forensic science service providers should express likelihood ratios as an order of magnitude or to one significant figure" (EWG 2024, pp. 82-84).

Table S2.9. NIST-extracted summary of responses provided in CTS Probabilistic Genotyping Test 23-5901/2 by 48 participants examining a 3-person mixture (based on mixing blood from three individuals in a 3:2:1 mixture). LR values come from comparison of the suspect DNA profile (Item 2) to the mixture DNA profile (Item 3) and are normalized to a common scientific notation format (e.g., 1E11 rather than 10^{11} or 100 billion). Information that is missing from the publicly available source is listed as (--). *also reported using ID Plus and MiniFiler along with the listed kit. LR = likelihood ratio; NoC = number of contributors; PGS = probabilistic genotyping software.

CTS WebCode	Item 4: Item 2	STR Kit	PGS	DNA mass conc (ng/µL)	DNA proportion (%)	NoC	LR Value for Item 2 in Item 4 (3:2:1)	Population Database
2DPRLB	Included	GlobalFiler	STRmix v2.7.0	--	59.00	3	3E27	NIST
2K2ZQM	Included	ESX 16 Fast	DNAxs	--	10.40	3	2.710E20	Local
2MZT8B	Included	GlobalFiler	STRmix v2.7.0	--	62.00	3	No LR "suspect used to condition"	--
362K2C	Included	PP21	STRmix v2.7	5.2500	54.00	3	4.6E10	Local
38V7WP	Included	Invest24plex	STRmix v2.5.11	0.5000	66.00	3	8E13	FBI
3HRA2H	Included	Fusion	STRmix v2.6.3	--	--	3	3.6E23, 6.7E26, 2.3E27	NIST
3XRZGZ	Included	Invest24plex	EuroForMix	1.0600	59.00	3	4.48E24	Local
4VGCFZ	Included	Fusion 6C*	STRmix	--	--	3	2.54E28	Local
6P8FA7	Included	ESI 17	LiRa v3.0	--	58.00	3	7.53E15	Local
6RADLM	Included	Fusion 6C	STRmix v2.6.3	--	--	3	8.0E27	NIST
6RRGV9	Included	GlobalFiler	STRmix	--	--	3	6.43E21	NIST
74PMLW	Included	GlobalFiler	STRmix v2.7.0	--	55.00	3	5.67E26	NIST
7LHKND	Included	Fusion	STRmix	--	--	3	2.0E13, 4.4E13, 1.1E14	NIST
82D8RW	Included	GlobalFiler	STRmix v2.7	--	52.00	3	7E12	NIST
82TQDH	Included	GlobalFiler	EuroForMix	--	--	3	6.70E36	Local
83Z7LA	Included	PP21	STRmix v2.8	0.1190	50.00	3	No LR (assumed contributor)	Local
8PEQHA	Included	PP21	STRmix v2.8	1.500	56.00	3	>1E11	Local
8WUCN8	Included	PP21	STRmix v2.8	0.0900	53.00	3	No LR (assumed contributor)	Local
B6H6QH	Included	Invest24plex	STRmix v2.5.11	0.4993	59.00	3	2E13	FBI

CTS WebCode	Item 4: Item 2	STR Kit	PGS	DNA mass conc (ng/µL)	DNA proportion (%)	NoC	LR Value for Item 2 in Item 4 (3:2:1)	Population Database
CJWUZ4	Included	PP21	STRmix v2.8	1.6080	59.00	3	No LR (assumed contributor)	Local
ERF8RY	Included	PP21	STRmix v2.8	1.7140	61.00	3	No LR (assumed contributor)	Local
G632HU	Included	GlobalFiler	STRmix v2.7	1.2040	62.00	3	No LR (assumed contributor)	NIST
GRCZ6X	Included	Fusion	TrueAllele	--	--	3	1.1E21, 2.5E24, 4.6E24	NIST
HDMM48	Included	Fusion 6C	STRmix v2.6.3	0.5310	9.00	3	6.8E28	NIST
J99CLX	Included	PP21	STRmix v2.8.0	--	60.68	3	No LR conditioned on Item 2	Local
JGHZZV	Included	GlobalFiler	STRmix v2.7	--	--	3	at least 1E11	Local
JQ7NY6	Included	ID Plus	EuroForMix	26.21	53.42	3	1.39E14, 1.51E14, 3.98E14	NIST
JX9X7X	Included	PP21	STRmix	0.2000	56.00	3	No LR (assumed contributor)	Local
JXTGB4	Included	Fusion	STRmix	--	--	3	5.6E11, 8.1E12, 8.4E12	NIST
K8XREJ	Included	Fusion*	STRmix	--	--	3	8.10E26	Local
KVY9FT	Included	Fusion 5C	STRmix	--	--	3	3.4E20, 6.9E23, 2.5E24	NIST
LPPCAY	Included	ESX 17 Fast	STRmix v2.5.11	0.8400	54.00	3	3.0E18	Local
NMAQDU	Included	PP21	STRmix v2.8	0.6351	58.00	3	No LR (assumed contributor)	Local
NN6EU4	Included	Invest24plex	STRmix v2.5.11	0.6140	59.00	3	3E13	FBI
P8FLKQ	Included	PP21	STRmix	0.8300	57.00	3	No LR (assumed contributor)	Local
QCP2RU	Included	Fusion	STRmix	--	--	3	5.9E21, 1.4E25, 5.3E25	NIST
QJA3AR	Included	PP21	STRmix v2.8	1.7530	61.00	3	1E11 (item 2 is assumed)	Local
RAEJ9N	Included	ID Plus	STRmix v2.7.0	--	--	3	7.09E13	NIST
T46M3T	Included	Fusion 5C	STRmix	--	--	3	2.7E22, 5.4E25, 2.2E26	NIST

CTS WebCode	Item 4: Item 2	STR Kit	PGS	DNA mass conc (ng/µL)	DNA proportion (%)	NoC	LR Value for Item 2 in Item 4 (3:2:1)	Population Database
THNB9G	Included	GlobalFiler	STRmix	--	52.60	3	--	FBI
TRB7DR	Included	Fusion 6C	STRmix	--	--	3	--	--
UHWYPN	Included	PP21	STRmix v2.8.0	0.9800	60.00	3	No LR (assumed contributor)	Local
UUZ3CG	Included	GlobalFiler	STRmix v2.7.0	--	56.00	3	2.83E23	NIST
YAB63G	Included	ID Plus	STRmix v2.7	--	--	3	1.46E14	NIST
YDBGXG	Included	ID Plus	STRmix v2.7.0	--	--	3	8.7E13	NIST
Z78W9Q	Included	NGM Detect	--	--	54.00	3	1.7E9	Local
Z8KMYL	Included	Fusion	TrueAllele	--	--	3	--	NIST
ZTYATR	Included	Invest24plex	STRmix	2.2400	57.00	3	9.64E19	NIST

In this second PGS PT, 48 participants utilized 10 different STR kits: PowerPlex 21 (12x), GlobalFiler (10x), PowerPlex Fusion/Fusion 5C (9x), Investigator 24plex (5x), PowerPlex Fusion 6C (4x), Identifiler Plus (4x), ESX 16 Fast (1x), ESX 17 Fast (1x), ESI 17 (1x), and NGM Detect (1x) along with five different PGS systems: STRmix (40 of 48; 83%), EuroForMix (3x), TrueAllele (2x), DNAxs (1x), and LiRa (1x) with one participant not reporting. At least four different versions of STRmix were used: v2.5.11 (4x), v2.6.3 (3x), v2.7 (11x), and v2.8 (9x) with 13 not listing any version number. All 48 participants selected NoC=3. Most participants in this PT either used a location identifying database (21 of 48; 44%) or the NIST population data (21 of 48; 44%) with the remainder using FBI population data (4 of 48; 8%) or not providing this information (2 of 48; 4%).

Assigned LR values ranged across 27 orders of magnitude from 10^{9} to 10^{36}. Almost a quarter of the participants (11 of 48; 23%) did not provide an LR value stating that the Item 2 (suspect) was an assumed contributor based on the case scenario provided, which was that the evidence (Item 4) was a stain from the pants that the suspect (Item 2) was wearing.

Table S2.10. NIST-extracted summary of responses provided in CTS Probabilistic Genotyping Test 23-5904/5 by 72 participants examining a 2-person mixture (based on mixing blood from two individuals in a 3:2 mixture). LR values come from comparison of the suspect DNA profile (Item 2) to the mixture DNA profile (Item 4) and are normalized to a common scientific notation format (e.g., 1E11 rather than 10^{11} or 100 billion). Information that is missing from the publicly available source is listed as (--). LR = likelihood ratio; NoC = number of contributors; PGS = probabilistic genotyping software.

CTS WebCode	Item 4: Item 2	STR Kit	PGS	DNA mass conc (ng/µL)	DNA proportion (%)	NoC	LR Value for Item 2 in Item 4 (3:2)	Population Database
2PM2BN	Included	ID Plus	STRmix	--	--	2	>1E12	NIST
2Z6NFH	Included	ID Plus	STRmix v2.7	--	--	2	3.30E17	Caucasian
34F2TB	Included	NGM Select	STRmix	--	--	**3**	1E9 (3.2E13)	Local
37BFNF	Included	PP21	STRmix v2.8	0.1290	100.00	2	1E11	Local
38MCUK	Included	ID Plus	STRmix v2.7	--	--	2	>1E12	NIST
3NPYGD	Included	Fusion 6C	STRmix v2.6.3	--	--	2	1.2E28	NIST
3UF9M6	Included	Fusion	STRmix	--	--	1(sp)/ 2(e)	RMP >7.2E9	NIST
3XH6EK	Included	ID Plus	STRmix v2.7	--	--	2	>1E12	NIST
3YA6CW	Included	GlobalFiler	STRmix	2.1200	100.00	1(sp)/ 2(e)	>1E9 (9.0498E25, 1.0535E26, 1.2223E28)	NIST
4KNADB	Included	Fusion 6C	STRmix v2.6.3	--	--	2	at least 1.3E28	NIST
4THF7C	Included	GlobalFiler	STRmix	0.2320	100.00	2	--	FBI-CAUC
64EZR2	Included	PP21	STRmix v2.7	11.98	100.00	2	5.4E9, >1E11	Local
6M9JZG	Included	ID Plus	STRmix v2.7	--	--	2	3.30E17	Caucasian
6MQMEA	Included	PP21	STRmix v2.8.0	0.2480	99.00	2	1E11	Local
6NGX63	Included	ESI 17	STRmix v2.7.0	--	--	2	at least 1E9	Local
6PXPHT	Included	GlobalFiler	STRmix	22.91	100.00	1(sp)/ 2(e)	>1E9 (9.0280E25, 1.0343E26, 1.0792E28)	NIST
6V2RPE	Included	PP21	STRmix v2.8	1.2140	88.00	**3**	1E11	Local
7669R6	Included	Invest24plex	EuroForMix	--	--	1(sp)/ 2(e)	1.19E25	Local

CTS WebCode	Item 4: Item 2	STR Kit	PGS	DNA mass conc (ng/µL)	DNA proportion (%)	NoC	LR Value for Item 2 in Item 4 (3:2)	Population Database
8KWN9A	Included	GlobalFiler	LRmix Studio	--	--	2	6.2E26	--
97RNXH	Included	GlobalFiler	STRmix v2.10	--	--	2	3.31E28	FBI
984GV7	Included	GlobalFiler	STRmix v2.6.3	0.3760	100.00	2	>1E11	Local
9FDPER	Included	ID Plus	STRmix v2.7	--	--	2	3.30E17	Caucasian
9ME3CA	Included	ID Plus	EuroForMix	42.51	100.00	2	RMP=1.89E18 5.17E14	NIST
9UUNG8	Included	GlobalFiler	STRmix	--	--	2	3.2319E25	FBI
A62BEE	Included	GlobalFiler	--	4616.0/ 1154.0	80.00/ 20.00	2	1.7533E20	Local
AWC7HB	Included	GlobalFiler	STRmix	--	78.74/ 21.26	2	3.55E28	FBI
B6LDTP	Included	ID Plus	STRmix v2.7	--	--	2	>1E12	NIST
BAG638	Included	ID Plus	STRmix v2.7	--	--	2	3.30E17	Caucasian
DP7WYT	Included	Fusion	STRmix	--	--	2	RMP >7.2E9	NIST
DW9WHV	Included	GlobalFiler	STRmix	--	--	1(sp)/ 2(e)	(sp) 1.6709E24 (e) 9.2358E20	--
EFDYPY	Included	Fusion 6C	STRmix v2.6.3	--	--	2	1.2E28	NIST
EXU6WN	Included	Fusion	STRmix	--	--	1	RMP >7.2E9	NIST
F3GTW8	Included	GlobalFiler	STRmix	--	--	2	3.5447E28	--
FBB2HX	Included	GlobalFiler	STRmix "version 10"	20.45/ 3.7500	84.49/ 15.51	2	2.09E28	Local
FYLKQW	Included	NGM Detect	EuroForMix	30.00/ 10.00	75.00/ 25.00	2	1E20	Local
GHT9V2	Included	PP21	STRmix v2.8	0.8080	100.00	1(sp)/ 2(e)	(sp) 1E11 (e) 1E11	Local
HGV4VY	Included	GlobalFiler	STRmix	--	--	2	2.6594E25	FBI
HQ4N32	Included	PP21	STRmix v2.8	0.0200	100.00	2	1E11	Local
JQHD28	Included	ID Plus	STRmix v2.7.0	--	--	2	>1E12	NIST
JRRPBW	Included	GlobalFiler	STRmix	--	--	2	3.1656E25	FBI
KL2R8R	Included	PP21	STRmix v2.8.0	0.0950	100.00	2	1E11	Local

CTS WebCode	Item 4: Item 2	STR Kit	PGS	DNA mass conc (ng/µL)	DNA proportion (%)	NoC	LR Value for Item 2 in Item 4 (3:2)	Population Database
LAL4PD	Included	ID Plus	STRmix v2.7	--	--	2	3.30E17	Caucasian
LK4N8P	Included	Fusion 6C	STRmix v2.6.1	--	--	2	--	--
M3EEQZ	Included	GlobalFiler	STRmix v2.10	--	81.39/ 18.61	2	1.69E28	FBI
M3K38Y	Included	PP21	STRmix v2.8	0.0100	98.00	2	1E11	Local
MGJ7CP	Included	PP21	STRmix v2.8	0.8600	100.00	1(sp)/ 2(e)	(sp) >1E11 (e) >1E11	Local
MTFYZF	Included	Fusion	STRmix	--	--	2	RMP >7.2E9	NIST
MX7EPM	Included	Invest24plex	STRmix v2.5.11	0.6599	100.00	2	at least 4E14	FBI
NAXHBT	Included	Fusion 6C	STRmix v2.6.3	--	--	2	at least 1.3E28	NIST
NERWYW	Included	GlobalFiler	STRmix	--	--	2	3.6728E28	Local
NNVP4N	Included	GlobalFiler	LRmix Studio	3.6100	100.00	2	7.6725E13	Local
NZLZFU	Included	ESX 16 Fast	DNAxs	--	--	2	5.651E19	Local
P847TY	Included	ID Plus	STRmix	--	--	2	>1E12	NIST
PAQQUW	Included	NGM Detect	EuroForMix	--	--	2	1E22	Local
PNAD4U	Included	Fusion 6C	LRmix Studio, Lab Retriever, DNA View, EuroForMix	--	--	2	LRmix Studio 6.6E27 Lab Retriever 6.7E24 DNA View 7.0E29 EuroForMix 2.7E29	NIST
QA94VE	Included	Fusion	STRmix v2.6.3	--	--	2	RMP >7.2E9	NIST
QP6EMQ	Included	ID Plus	STRmix v2.7.0	--	--	1(sp)/ 2(e)	>1E12	NIST
R8YCLX	Included	ID Plus	STRmix	--	--	2	>1E12	NIST
RVACVV	Included	ID Plus	STRmix v2.7	--	--	Two	>1E12	NIST
T9TEQR	Included	PP21	STRmix v2.8.0	0.2000	100.00	2	1E11	Local
TBJGXB	Included	Fusion 6C	STRmix v2.6.2	--	100.00	2	1E28	NIST

CTS WebCode	Item 4: Item 2	STR Kit	PGS	DNA mass conc (ng/µL)	DNA proportion (%)	NoC	LR Value for Item 2 in Item 4 (3:2)	Population Database
TXY2UB	Included	Fusion 6C	LRmix Studio	716.0	100.00	2	1.36E21	STRidER
VKR849	Included	Fusion	STRmix v2.6.3	--	--	2	RMP >7.2E9	NIST
VLQ8CL	Included	PP21	STRmix	0.8907	100.00	1(sp)/ 2(e)	(sp) >1E11 (e) >1E11	Local
VM2AYK	Included	PP21	STRmix v2.8	1.5720	100.00	2	(sp) 1E11 (e) 1E11	Local
VMFH4E	Included	Invest24plex	STRmix v2.5.11	0.8460	100.00	1(sp)/ 2(e)	(sp) at least 4E14 (e) at least 1E14	FBI
WJN2WU	Included	GlobalFiler, ESI Fast	--	--	--	2	RMP >1E9	Local
WKY4HR	Included	GlobalFiler	STRmix v2.5.11	--	--	2	6.14147E14	ABI Caucasian
WMBL89	Included	Fusion	STRmix	--	--	1(sp)/ 2(e)	RMP >7.2E9	NIST
WXUXJL	Included	GlobalFiler	STRmix	--	--	2	3.3369E25	FBI
XU9A6L	Included	GlobalFiler	STRmix	--	--	2	9.12E26	--
YF73QG	Included	GlobalFiler	STRmix v2.10	--	82.19/ 17.81	2	3.63E28	Local

In this third PGS PT, 72 participants utilized 11 STR kits: GlobalFiler (22x), Identifiler Plus (15x), PowerPlex 21 (12x), PowerPlex Fusion 6C (8x), PowerPlex Fusion (7x), Investigator 24plex (3x), NGM Detect (2x), NGM Select (1x), ESI 17 (1x), ESI Fast (1x), and ESX 16 Fast (1x) along with six PGS systems: STRmix (61 of 72; 85%), EuroForMix (5x), LRmix Studio (4x), DNA View (1x), DNAxs (1x), and Lab Retriever (1x) with two participants not reporting. At least seven different versions of STRmix were used: v2.5.11 (3x), v2.6.1 (1x), v2.6.2 (1x), v2.6.3 (7x), v2.7 (13x), v2.8 (10x), and v2.10 (4x) with 22 not listing any version number. Almost all participants (70 of 72; 97%) reported NoC=2 or one with the sperm fraction and two with the epithelial fraction (i.e., "1(sp)/2(e)"), which multiple participants did (11 of 72; 15%). The remaining two participants (2 of 72; 3%), 34F2TB and 6V2RPE, reported NoC=3. When recorded[22], most of the DNA concentration and proportion information comes from Table 6 of the CTS summary report under Item 4sp (Item 4 sperm fraction) following differential extraction. Again, most participants in this PT either used a location identifying database (25 of 72; 35%) or the NIST population data (25 of 72; 35%) with the remainder using FBI population data (10 of 72; 14%), "Caucasian" (6 of 72; 8%), STRidER (1 of 72; 1%), or not providing any information (5 of 72; 7%).

[22] Laboratory policies may influence some of the responses or lack thereof. For example, in the additional comments section of the CTS PT summary (Table 9 of Test 23-5904/5; see https://cts-forensics.com/reports/23-5904.5_Web.pdf), participant 9UUNG8 shared: "Reported/listed alleles for evidence samples include stutter peaks. Our laboratory does not report DNA concentration or proportion."

Assigned LR values ranged across 20 orders of magnitude from 10^9 to 10^{29}. As summarized in Table S2.10, participant PNAD4U shared LR values from 4 PGS systems for an effectively single-source profile following differential extraction of the Item 4 two-person mixture. These intra-laboratory assigned LRs ranged from 10^{24} to 10^{29}, with the lower LR values, as expected, coming from semi-continuous PGS systems that consider only alleles and not the peak heights (i.e., Lab Retriever and LRmix Studio). In addition, participant DW9WHV provided LR values for the two-person mixture in the "epithelial" fraction as 9.2×10^{20} and 1.67×10^{24} for the effective single-source "sperm" fraction, which illustrates an increase in the LR value that can be seen when the uncertainty decreases with the number of possible genotype combinations.

Table S2.11. NIST-extracted summary of responses to a 4-person mixture (prepared by mixing equal parts of blood from four individuals seeking to create 1:1:1:1) provided in CTS Probabilistic Genotyping Test 24-5901/2 by 62 participants. LR values come from comparison of the suspect DNA profile (Item 2) to the mixture DNA profile (Item 3). LR = likelihood ratio; NoC = number of contributors; PGS = probabilistic genotyping software.

CTS WebCode	Item 3: Item 2	STR Kit	PGS	DNA mass conc (ng/µL)	DNA proportion (%)	NoC	LR Value for Item 2 in Item 3 (1:1:1:1)	Population Database
27J4KB	Included	ESI17	LiRa v3.0	--	--	4	--	--
2TZLR8	Included	ID Plus	STRmix v2.5.11	0.01960	5.00	5	--	Local, NIST
3NJE48	Included	Fusion 6C	STRmix v2.5.11	--	7.00	4	1E23	NIST
4KE6C3	Included	PP21	STRmix v2.8	0.00200	12.00	4	5.3E9	Local
4Q2XN3	Excluded	PP21	STRmix	--	--	**3**	"The LR supports Hd"	Local
63EHK4	Included	GlobalFiler	STRmix v2.9.1	0.1900	2.40	4	3E7	NIST
6VPZR3	Included	GlobalFiler	STRmix	--	3.00	4	1.53E9	NIST
6XXAW4	Excluded	PP21	STRmix v2.8	--	--	**3**	"LR below 1"	Local
7T4FY3	Included	GlobalFiler	STRmix	0.1500	4.89	4	>1E9	NIST
9WGZW3	Included	ID Plus, Fusion 6C	STRmix v2.5.11	--	--	--	4.95E24	Local
9WVL74	Included	GlobalFiler	STRmix	0.1400	10.29	4	--	FBI Caucasian
A66AJ3	Included	GlobalFiler	STRmix v2.4	--	2.00	4	1.8E2 (SE33 & D1 not used)	NIST
AUMTRW	Included	PP21	STRmix v2.8	0.00035	7.00	4	1.9E9	Local
B7LWGW	No Interpretation	GlobalFiler	STRmix	--	--	5	--	--
C3L9DV	Included	GlobalFiler	STRmix	--	6.00	4	5.58E13	NIST
CKAFMX	Included	ID Plus	STRmix v2.7	--	--	--	6.6E11	NIST
CNQNBW	Included	ESX17 Fast	STRmix v2.5.11	--	--	4	--	--
CV4WCY	Included	GlobalFiler	LRmix Studio v2.1.3	--	--	--	2.4756E12	STRidER
CX6ZPV	Included	Fusion 6C	STRmix	--	6.00	4	1.83E28	Local
D4G86P	Included	PP21	STRmix	--	--	--	9.6674E11	PP21 stratified

CTS WebCode	Item 3: Item 2	STR Kit	PGS	DNA mass conc (ng/µL)	DNA proportion (%)	NoC	LR Value for Item 2 in Item 3 (1:1:1:1)	Population Database
E6NMCV	Included	ID Plus	STRmix	--	--	--	4.5E10	NIST
EDLFKP	Included	PP21	STRmix v2.8.0	0.00520	8.00	4	6.1E6	Local
G396TQ	Included	GlobalFiler	STRmix	--	5.00	4	4.35E15	NIST
GXTW6R	Included	GlobalFiler	STRmix v2.8	--	--	--	>1E12	FBI
GXXCQQ	Included	PP21	STRmix v2.8.0	0.00090	6.00	4	1E10	Local
HDQ4BR	No Interpretation	Invest24plex	--	--	--	4	--	--
J7339J	Included	PP21	STRmix	0.1753	6.87	4	>1E11	PP21
JGKK2Q	Included	ID Plus	STRmix	--	--	--	3.3E6	NIST
JJ74TP	Included	GlobalFiler	STRmix v2.8	--	--	--	>1E12	FBI
L26XRK	Inconclusive	PP21	STRmix v2.8	--	--	**3**	****	Local
L7CNVQ	Included	GlobalFiler	STRmix v2.4	--	4.00	4	1.5E18 (SE33 & D1 not used)	NIST
LE6Y6L	Included	PP21	STRmix v2.8	0.00100	8.00	4	2.4E10	Local
LFLWXL	Included	PP21	STRmix v2.8	0.00070	6.00	4	1E11	Local
LJL8TL	Included	PP21	STRmix	0.00072	12.00	4	1E11	Local
LRPGKK	Included	GlobalFiler	STRmix v2.8	--	--	--	>1E12	FBI
LWPP4K	Included	PP21	STRmix v2.8	0.00198	9.00	4	1E11	Local
NH4WTJ	Included	PP21	STRmix v2.8	0.00088	11.00	4	8.1E9	Local
NWU3LM	Inconclusive	Fusion 6C	--	--	--	4	--	--
P4RC6K	Included	GlobalFiler	STRmix	--	6.00	4	3.68E18	NIST
P9HLFJ	Included	GlobalFiler	STRmix v2.8	--	--	--	1.5E3	FBI
PKP22G	Included	PP21	STRmix v2.8	0.00099	9.00	4	1E11	Local
PXPZND	Excluded	PP21	STRmix	--	--	**3**	"Supports Exclusion"	Local

CTS WebCode	Item 3: Item 2	STR Kit	PGS	DNA mass conc (ng/µL)	DNA proportion (%)	NoC	LR Value for Item 2 in Item 3 (1:1:1:1)	Population Database
Q7ZLBM	Included	Fusion 6C, GlobalFiler, Precision ID GlobalFiler NGS v2	LRmix Studio, DNAxs, EuroForMix	0.01600	7.00	4	4.6E7 9.78E8 8.2E14 4.3E8 1.3E15 1.979E5	STRidER Europe
QETUVD	Included	PP21	STRmix v2.8.0	0.00162	9.00	4	1E11	Local
QGRATG	Excluded	GlobalFiler	--	--	--	4	--	Local
QKK2DL	Included	GlobalFiler IQC	STRmix v2.8.0	0.07600	6.66	4	1.4039E26	--
T9QWLD	Included	GlobalFiler	STRmix	--	6.00	4	6.48E21	NIST
U8YVVF	No Interpretation	Invest24plex	--	--	--	5	--	--
UAQU9D	Included	PP21	STRmix	0.00075	15.00	4	1.5E3	Local
V9EEEC	Included	PP21	STRmix v2.8.0	--	--	4	3.6E7	Local
VGN7HF	Excluded	Fusion 6C	DNAxs	--	--	4	--	--
VPY84F	Inconclusive	ID Plus	STRmix v2.7.0	--	--	4	--	--
VT27E9	Excluded	PP21	STRmix v2.8	--	--	**3**	"excluded"	Local
XHPQK7	Included	GlobalFiler	STRmix v2.9.1	--	--	5*	8.63E10	Local (stratified)
XJ2PFE	Included	Fusion 6C	STRmix v2.6.3	--	--	4*	>3.4E18	NIST
XKRQ7D	Included	GlobalFiler	STRmix	--	6.00	4	2.34E19	NIST
XNR32D	Included	GlobalFiler	STRmix	--	6.00	4	2.08E20	NIST
XUHAYB	No Interpretation	Invest24plex	EuroForMix	--	--	4	--	--
XWPRX8	Included	PP21	STRmix v2.10	0.4780	7.96	4	1.3E6	Local
Y43ZY9	Excluded	PP21	STRmix	--	--	**3**	--	Local
Y87VHC	No Interpretation	ID Plus	STRmix v2.7.0	--	--	**3**	-- "not suitable for comparison"	--
ZY2GWB	Included	GlobalFiler	STRmix	--	6.00	4	1.26E18	NIST

*The number of contributors for participants XHPQK7 and XJ2PFE are not cited in the Table 6, Item 3 Results section (p. 84 in CTS 24-5901/2), but are mentioned in the Table 7, Item 3 Methods and Results section (p. 100 in CTS 24-5901/2) as "considered to be a 5 person mixture" (XHPQK7) and "interpreted as a mixture of four individuals, the victim as an assumed contributor" (XJ2PFE)
**** Participant L26XRK declared: "The DNA evidence supports the exclusion of Item 2 from the DNA detected in item 3" (see p. 97 in CTS 24-5901/2) yet reported it as "Inc" (inconclusive; see p. 17 in CTS 24-5901/2).

In this fourth PGS PT, 62 participants utilized eight STR kits: GlobalFiler (23x), PowerPlex 21 (22x), PowerPlex Fusion 6C (7x), Identifiler Plus (7x), Investigator 24plex (3x), ESI 17 (1x), ESX 17 Fast (1x), and Precision ID GlobalFiler NGS (1x) along with five PGS systems: STRmix (53 of 62; 85%), EuroForMix (2x), LRmix Studio (2x), DNAxs (2x), and LiRa (1x) with four participants not reporting. At least seven different versions of STRmix were used: v2.4 (2x), v2.5.11 (4x), v2.6.3 (1x), v2.7 (3x), v2.8 (19x), v2.9.1 (2x), and v2.10 (1x) with 21 not listing any version number. Based on the STR kits being used, many of the PT results from this Test 24-5901/2 are not being performed by U.S. forensic DNA analysts since PowerPlex 21 and Identifiler Plus do not contain all of the required 20 NDIS STR loci.

Across the 62 participants, 41 (66%) selected NoC=4, 4 (6%) selected NoC=5, 7 (11%) selected NoC=3, and 10 (16%) did not provide an estimated NoC as captured in the CTS summary report. In some cases where no interpretation was made, it was due to the NoC determination. For example, notes provided in the summary report suggest that the participant was limited by their protocol to NoC=4 (e.g., participant B7LWGW provided a determination of NOC=5).

With this four-person mixture data summarized in Table S2.11 from 62 participants, there were seven false exclusions, three inconclusives, and five no interpretations. Importantly, there were no false inclusions observed. Participants providing the seven exclusionary decisions shared their comparison of Item 3 (mixture) to Item 2 (suspect) in different ways:

- "The LR supports Hd" (participant 4Q2XN3-5901; see Table 7, p. 93, in CTS summary)
- "LR below 1 (supports exclusion)" (participant 6XXAW4-5901; see Table 7, p. 94)
- "Supports exclusion" (participant PXPZND-5902; see Table 7, p. 98)
- "Excluded" (participant VT27E9-5902; see Table 7, p. 100)
- No comment on the Item 2 comparison (participant QGRATG-5901; see Table 7, p. 99; see also participant VGN7HF-5902; see Table 7, p. 100)
- Did not state anything about the LR value assigned but rather "The DNA evidence is more likely if the DNA originated from 3 unknown and unrelated people selected randomly from the [Location Identifying Population], than if it originated from the suspect and 2 unknown and unrelated people from the [Location Identifying Population]" (participant Y43ZY9-5902; see Table 7, p. 101).

Almost a quarter of the participants (14 of 62; 23%) did not report an LR value. The 48 participants who assigned LRs provided values ranging across 26 orders of magnitude from 10^2 to 10^{28}. Many of the false exclusions come from participants who made a determination of NoC=3 rather than the correct NoC=4.

6.4. **Comments on Publicly Accessible Proficiency Test Results**

Proficiency test data, although not always available with more than two person mixtures, provides an opportunity to observe performance across the entire process. There appears to be more variation when three- or four-person mixtures are examined in exercises with publicly available PT results. Commenting on what can be learned from proficiency testing information (based on an earlier version of Table S2.5), a group of four researchers and practitioners stated following their own review of CTS PT data:

> "None of the false positives or negatives could be attributed to the mixture interpretation strategy and certainly not to the use of PGS…. In the end, proficiency test data are currently not a good metric to judge the overall reliability of a system. Individual laboratory systems can use the results to determine how the individual participants performed since the labs know the conditions and parameters of their analysis and reporting. In addition, there are no restrictions on who can participate in vendor-provided proficiency tests, meaning these tests can be used for training, research, or academic purposes. Attempting to judge the overall reliability of a discipline/system using proficiency test data without knowing the sources and causes of each discordant result is misleading and uninformative.… If proficiency test data are going to be used to evaluate reliability, a more in-depth examination must be performed" (Bille et al. 2022).

Three different types of PTs have been examined here based on their sample source or requested output: (1) those providing single-source or simple two-person mixtures as biological samples in the form of blood and/or semen (Table S2.5) that permit assessing the entire measurement and interpretation process, (2) those providing single-source, two-person, or three-person mixtures as DNA profile electropherograms (Table S2.6) that permit assessing only the interpretation process, and (3) those providing two-, three-, and four-person mixtures as biological samples (Table S2.7) where a PGS output with assigned LR values can be reported (Tables S2.8 to S2.11) that permit assessing the entire DNA mixture measurement and interpretation process.

Collectively across the 135 PTs examined, each with two evidence items (i.e., test samples "Item 3" or "Item 4", so 270 potential mixtures), there were 42% single-source samples (114 of 270), 52% two-person mixtures (140 of 270), 5.6% three-person mixtures (15 of 270), and 0.4% four-person mixtures (1 of 270). Thus, most of these 135 CTS DNA mixture PTs summarized in Tables S2.5 to S2.11 involve single-source samples, or two-person mixtures created from large quantities of DNA (hundreds to thousands of cells) from semen and/or blood. This is unlike typical casework performed today involving complex mixtures (i.e., >2p mixtures often with allele drop-out from low quantity contributors).

A majority of the mixtures involving biological samples were created with blood and semen (83 of 117; 71%) compared to mixtures with multiple blood samples (34 of 117; 29%). These samples were typically combined in approximately one-to-one (1:1) ratios. In other words, the mixtures in the Forensic Biology, DNA Semen, and DNA Mixture PT exams (Table S2.5) are not complex. Therefore, with this PT data we cannot assess how DNA analysts may or may not perform with three-, four-, or five-person mixtures. It may be worth noting as well that with

differential extraction procedures enabling chemical separation of sperm from epithelial cells, many of the comparisons and interpretations would effectively be to single-source male DNA profiles rather than to profiles produced from a mixture from blood and semen.

Information regarding inclusion or exclusion of a reference sample (Item 1 or Item 2) to the mock evidence sample (Item 3 or Item 4) may assist participants in seeing how they performed relative to others in a particular proficiency test. However, this information alone is not helpful in understanding factors that may influence inconclusive decisions or variability in LR assignments. In the future, it may be helpful for PT providers to collect more information and details with participant responses. Forensic DNA laboratory case reports involving DNA mixture interpretation – especially when PGS is used – typically describe findings with numerical LR values, sometimes along with verbal equivalents (SWGDAM 2018). When no LR values are shared by participants (e.g., in the biological sample PTs or the DNA Interpretation PTs), those who wish to assess this information to explore proficiency of the participants are unable to do so.

Approximately 6% of CTS PT data involve DNA mixtures with more than two contributors. In the next sections, some observed differences with two-person versus three-person mixtures are explored followed by a closer look at two PTs involving three-person (3p) and four-person (4p) mixtures with some low-quantity DNA contributors.

6.4.1. **Observed Differences with Two-Person versus Three-Person Mixtures**

The DNA Interpretation PTs summarized in Table S2.6 provide a small window into some performance differences with 2p and 3p mixtures. In total, these 22 datasets provide 2520 responses (630 participants × two evidence items × two reference items) with three (0.12%) false inclusions, 15 (0.60%) false exclusions, 156 (6.2%) inconclusive results, and 10 (0.40%) no responses. The three false inclusions came from 2p mixtures where the reference profile of interest was not present in the provided evidence item. All 15 false exclusions came with 3p mixtures. With the 156 inconclusive results, 15 (9.6%) came from 2p mixtures and 141 (90.4%) came from 3p mixtures. For the 10 “no response” decisions, six (60%) came from 2p mixtures and four (40%) came from 3p mixtures. Thus, in this dataset, all 15 false exclusions came with 3p mixtures as well as over 90% of the inconclusive decisions.

In the 15 false exclusions, 13 (87%) came from a single three-contributor mixture that was part of a DNA Interpretation proficiency test conducted at the end of 2022 (CTS 22-5882). Among 35 responses provided in this test for Item 3, which was prepared as a 3:1:1 mixture (with no DNA quantity reported) and included the suspect (Item 2) as a minor contributor, 11 participants correctly included the suspect, 13 incorrectly excluded, and 11 provided an inconclusive response. *Thus, with three possible responses, roughly a third of participants correctly included the suspect, a third incorrectly excluded him, and a third made an inconclusive decision.*

6.4.2. **A Three-Person Mixture in CTS DNA Interpretation Test 22-5882**

In Table S2.6, the second CTS DNA Interpretation Test from 2022 (22-5882) involved 35 participants. For the Item 4 portion of the test, which used a 2p mixture in a 1:3 ratio, all

participants correctly placed the Item 2 donor in the mixture and correctly excluded the Item 1 non-donor. However, with the Item 3 portion of the test, which involved a 3p mixture in a 1:3:1 ratio with a low amount of DNA from the contributor of interest (Item 2), there was a disparity in the responses for the Item 2 suspect with 11 (31%) including, 13 (37%) excluding, and 11 (31%) reporting inconclusive. The correct answer was to include the Item 2 suspect in this 3p mixture. An examination of the reported consensus STR profile found 17 alleles from Item 2 missing in the Item 3 three-person mixture's consensus STR profile (see p. 3 in CTS 22-5882) presumably from stochastic effects with amplifying low-levels of DNA. Like regular casework involving complex DNA mixtures, allele drop-out is to be expected due to stochastic effects from PCR amplification of contributor(s) containing low-quantities of DNA.

The same participants in CTS 22-5882 had no difficulties with correctly excluding Item 1 (the victim) and including Item 2 (the suspect) in Item 4, which was a two-person 3:1 mixture that exhibited no allele drop-out in the consensus profile. However, the LR values reported by participants (see Table 6 on pp. 61-64 in CTS 22-5882) ranged from greater than 10^6 ("actual LR available upon request") to 10^{53}, or 47-orders of magnitude. Some of this variation on the lower end of assigned LR values may be explained by the use of reporting caps (i.e., $>10^6$ with actual LR available upon request). Since LR values are not required in the DNA Interpretation PT, only a fraction (9 of 35) of participants reported an assigned LR value for Item 3 while 25 of 35 reported an assigned LR value for Item 4.

6.4.3. **A Four-Person Mixture in CTS Probabilistic Genotyping Test 24-5901/2**

With the four PGS PTs that CTS has provided since 2022, participants have been asked to report on comparisons with four 2p, three 3p, and one 4p mixtures (see Table S2.7). To assess the degree of reliability with PGS results, examining variation in assigned LR values across test participants is important beyond the current categorical responses of inclusion, exclusion, inconclusive, or no response. The NIST-extracted summaries in Tables S2.8 to S2.11 endeavored to do this. However, details including assumptions made and propositions used are often lacking to prevent a full understand of the variability that is observed in the assigned LR values. This is especially so in a four-person mixture assessed by 62 participants in a 2024 PGS proficiency test (CTS 24-5901/2).

Participants were all provided the same biological samples either on cloth swatches (test version 5901) or FTA micro cards (test version 5902). This PT enabled participants to perform serological screening for blood, saliva, and semen and to report whether the stain was of human origin and contained any male donors via a Y-chromosome assay. Participants conducted laboratory-specific DNA extraction, DNA quantitation, PCR amplification using various STR typing kits and conditions, interpreted the resulting STR profiles with various PGS systems, and reported assigned LRs using various allele frequency population databases (see Table S2.11).

The manufacturer's information provided details for Item 3, which was a four-person mixture spotted on colored fabric: "Item 3 was created by combining one part blood from the Item 1 female donor, one part blood from the Item 2 male donor, and one part blood each from another female and male donor whose known standards were not provided" (see p. 2 in CTS

24-5901/2). For verification purposes, samples were sent out to several predistribution sites prior to release of the PT materials. CTS noted following this verification step: "Consistent allelic results were reported for all STR loci across both substrates, with the exception of Item 3. Predistribution participants were missing one or more alleles at a few loci" (see p. 2 in CTS 24-5901/2). In their summary comments (revised July 24, 2024), CTS reported:

> "Item 3 was created with blood from four donors, two females and two males. A significant number of participants were missing alleles compared to the Manufacturer's preparation information[23]…Of the 62 participants that reported STR results for this item, only 14 (23%) reported full allelic results that were consistent with one another and the Manufacturer's preparation information. Of the remaining participants, 33 (53%) were missing alleles that could be attributed to the Item 2 male donor and 15 (24%) were missing alleles that could be attributed to two or more donors…" (see p. 4 in CTS 24-5901/2).

Thus, for many of the participants, some of the contributors in this complex mixture contained low-quantity DNA, perhaps due to poor DNA extraction or PCR amplification (e.g., if there was an inhibitor in the colored fabric). *This type of complex mixture with some low-quantity DNA contributors exhibiting allele drop-out is what forensic DNA laboratories commonly face with DNA mixture interpretation for routine casework.* Differences observed might also come from variability in sample preparation by the PT provider or evidence sampling by the PT participant.

As described in the CTS summary report and summarized in Table S2.11, the selected number of contributors for the 62 participants varied from NoC=3 (7 of 62; 11%), NoC=4 (41 of 62; 66%), NoC=5 (4 of 62; 6%), or did not provide an estimated NoC (10 of 62; 16%). With the 48 participants who assigned LRs using various PGS systems, responses ranged across 26 orders of magnitude from 10^2 to 10^{28} along with seven false exclusions (i.e., LR<1), many of which came from participants that made an incorrect determination of NoC=3 rather than the correct NoC=4. There were no false inclusions reported.

In the four PGS PTs, reporting formats varied as did the propositions used and range of assigned LR values reported, which was typically more than 20 orders of magnitude on the same provided DNA sample. More details on DNA extraction and PCR amplification protocols as well as propositions used for LR assignment would be helpful. Additional experience with these types of PT results may assist with improved understanding of DNA mixture interpretation variability and the degree of reliability of PGS systems for the factors explored in these studies.

[23] There were seven autosomal STR markers (D7S820, D12S391, D13S317, D19S433, D21S11, Penta D, Penta E) and two Y-chromosome STR markers (DYS570 and DYS576) where results were not received by a minimum of 10 participants (see p. 3 in CTS 24-5901/2).

7. Interlaboratory Studies Involving DNA Mixture Interpretation

Interlaboratory comparison studies, which are sometimes referred to as collaborative exercises or round-robin studies, provide a community-focused approach to assess whether multiple analysts and laboratories can generate comparable measurements and interpretation when provided with the same samples or DNA profiles. Ideally, we would like to characterize uncertainty for the entire system (mixture sample to reported result), but interlaboratory studies, like proficiency tests discussed in the previous section, typically only provide a partial picture of the variability because of the difficulty of creating and providing consistent mixture samples to many participants.

It is important to keep in mind that interlaboratory studies, which are typically published in peer-reviewed literature, assess variability in practice at the time they are conducted. Participants are volunteers who share their time and resources with support of their laboratory management as they examine the various DNA mixtures designed for study.

The DNA Commission of the International Society for Forensic Genetics (ISFG) stated in their 2006 “Recommendations on the interpretation of mixtures” article that “our discussions have highlighted a significant need for continuing education and research into this area" (Gill et al. 2006). Interlaboratory studies provide both. The research conducted explores variability across the community of participants, and the results help educate participating analysts and laboratories regarding their performance relative to others.

There have been at least 20 interlaboratory studies involving DNA mixture interpretation (Table S2.12). These studies have been organized by NIST, the Defense Forensic Science Center (DFSC), the Spanish-Portuguese Working Group of the International Society for Forensic Genetics (GHEP-ISFG), the European Forensic Genetics Network of Excellence (EuroForGen-NoE), the Netherlands Forensic Institute (NFI), developers of the PGS system STRmix (the Institute of Environmental Science and Research Limited, ESR), the UK Forensic Science Regulator, the UK Association of Forensic Science Providers (AFSP), and most recently by Noblis and Bode Technology via National Institute of Justice (NIJ) funding (Brinkac et al. 2023, Hicklin et al. 2023a, Hicklin et al. 2023b) and researchers from Sam Houston State University, ESR, and the University of Auckland (Boodoosingh et al. 2024).

The NIST-extracted summary of information found in Table S2.12 has been ordered chronologically with the year(s) of the study on the left side followed by the publication citation and a name for the study. When a PGS system was utilized by participants, this information is indicated along with the software version, if described. Where no PGS system was used or available for DNA mixture interpretation, such as prior to 2013, entries are marked “N/A” (not applicable). Where details were unavailable in publicly accessible information, such as with the total DNA amount (e.g., Butler et al. 2018, Crespillo et al. 2014), entries are marked “N.E.S.” (not explicitly stated).

Columns in Table S2.12 summarize the samples or data provided to participants and in what format, the number of participating laboratories and data sets received, the number of samples provided in the study with a breakdown by the number of contributors, the total DNA amount or range of amounts in the various samples assessed, and the range of mixture ratios explored.

Table S2.12. NIST-extracted summary of information from 20 interlaboratory studies involving DNA mixture interpretation. Abbreviations: 2p = two-person mixture; 3p = three-person mixture; 4p = four-person mixture; 5p = five-person mixture; AT = analytical threshold; N/A = not applicable; N.E.S. = not explicitly stated; NoC = number of contributors; pg = picograms; ss = single-source; S&S = Schleicher & Schuell; Unk. = unknown; Year = year study was conducted.

Year	Reference & Study Name	PGS System (Version)	Format of Sample/Data Provided	# Labs (Data Sets)	# Sam-ples	# with NoC	Total DNA Amount (pg)	Mixture Ratio Range
1997	Duewer et al. (2001) NIST Mixed Stain Study #1	N/A	Buffy coat cells on S&S 903 paper	22 (37)	11	6-ss 4-2p 1-3p	30,000 to 50,000 30,000 to 50,000 30,000 to 50,000	N/A ≈1:1 ≈1:1:1
1999	Kline et al. (1999); Duewer et al. (2001) NIST Mixed Stain Study #2	N/A	Blood & semen stains on cotton cloth; DNA extracts	45 (70)	11	4-ss 6-2p 1-3p	≈1 µg per source, or ≈1 to 3 million pg for each stain; 500 to 5,000 pg/µL for DNA extracts	3:1 2:1:1
2001	Kline et al. (2003); Duewer et al. (2004) NIST Mixed Stain Study #3	N/A	DNA extracts	74 (117)	6	1-ss 5-2p 1-3p	1,000 to 4,000 pg/µL	3:1 to 10:1 4:2:1
2005	Butler et al. (2018) NIST MIX05	N/A	EPG data (.fsa files) from 6 STR kits	69 (75)	4	4-2p	N.E.S.	1:1 to 7:1
2010	Crespillo et al. (2014) GHEP-MIX01	N/A	EPG data (.fsa files) from 2 STR kits	32 (32)	4	4-2p	N.E.S.	1:1 to 10:1
2011	Crespillo et al. (2014) GHEP-MIX02	N/A	EPG data (.fsa files) from 1 STR kit	24 (24)	2	1-2p 1-3p	N.E.S.	5:1 2:1:1
2012	Crespillo et al. (2014) GHEP-MIX03	N/A	EPG data (.fsa files) from 2 STR kits	17 (17)	3	2-2p 1-3p	N.E.S.	5:1 to 10:1 7:3:1
2013	Prieto et al. (2014) EuroForGen Mixture Study	LRmix by all labs	EPG data (csv format) with case scenarios; population allele frequencies	18 (20); 18 (22)	2	2-2p	N.E.S.	N.E.S.

Year	Reference & Study Name	PGS System (Version)	Format of Sample/Data Provided	# Labs (Data Sets)	# Sam-ples	# with NoC	Total DNA Amount (pg)	Mixture Ratio Range
2013	Butler et al. (2018) NIST MIX13	Lab Retriever or TrueAllele used by 3 labs	EPG data (.fsa files) from 2 STR kits with case scenarios	108 (163)	5	2-2p 2-3p 1-4p	N.E.S.	1:1 to 3.5:1 6:1.5:1, 7:2:1 1:1:1:1
2014	Barber et al. (2015) UK Forensic Regulator Mallinder et al. (2022)	LRmix, likeLTD used by 2 labs	4 DNA mixtures and 1 EPG data (.fsa file) with case scenarios	8 (18)	5	2-2p 3-3p	N.E.S.	2:1 to 4:1 6:4:1 to 7:1.5:1
2014-2015	Aranda et al. (2015) *talk* DFSC Mixture Study Rogers et al. (2022)	N.E.S.	N.E.S.	55 (185)	6	4-2p 2-3p	N.E.S.	2:1 to 3.5:1 1:1:1, 4:1:1
2014	Cooper et al. (2015) STRmix	STRmix (2.0?) by all labs	Identifiler profiles from 3 casework samples (ground truth not known)	12 (20)	3	Unk.	N.E.S.	Unk.
2014	Toscanini et al. (2016) GHEP-ISFG Basic	N/A	Stain from 2:1 volume ratio mixture of saliva and blood	72	1	1-2p	N.E.S.	≈2:1
2014	Toscanini et al. (2016) GHEP-ISFG Advanced	N/A	Stain from 4:1 volume ratio mixture of saliva and semen	52	1	1-2p	N.E.S.	≈4:1
2015	Barrio et al. (2018) GHEP-ISFG MIX06	LRmix Studio used by 15 labs	EPG data (PDF) for NGM kit loci pre-analyzed with AT = 50 RFU	25	1[§]	1-3p	N.E.S.	7:3:1

Year	Reference & Study Name	PGS System (Version)	Format of Sample/Data Provided	# Labs (Data Sets)	# Sam-ples	# with NoC	Total DNA Amount (pg)	Mixture Ratio Range
2016	Benschop et al. (2017a) NFI-organized inter- and intra-laboratory exercise	LRmix Studio (v2.0.1) used by 1 lab on some samples	EPG data (PDF) with 4 replicates for NGM kit loci pre-analyzed with AT = 50 RFU; provided in Sets A or B	3 (26)	5 in each of 2 sets	2-2p 4-3p 2-4p 2-5p	180 24 27 186 360 240 1750	5:1 1:1 1:1:1 25:5:1 10:1:1 5:1:1:1 2:2:1:1:1
2018	Thomson (2018) *talk* UK AFSP Mallinder et al. (2022)	5 STRmix, 1 LiRa, 1 LRmix/ EuroForMix	Re-used DNA mixtures from Barber et al. (2015)	7 (28)	4	2-2p 3-3p	N.E.S.	2:1 to 4:1 6:4:1 to 7:1.5:1
2018	Bright et al. (2019a) STRmix collaborative exercise	STRmix (v2.4 and v2.5)	2 PROVEDIt EPG profiles (.hid files) or text files with STR allele, peak height, and size information; *key known variables were held constant*	42 (174)	2	1-3p 1-4p	750 105	4:4:1 4:1:1:1
2021-2022	Hicklin et al. 2023a DNAmix 2021 NoC	N/A	EPG profiles (.hid files)	67 (134)	21	2-2p 9-3p 8-4p 1-5p 1-6p	43 to 872	≈1:1, 2:1 ≈1:1:1 to 3:2:1 ≈1:1:1:1 to 2:2:2:1 ≈2:1:1:1:1 ≈3:3:3:2:2:1
2021-2022	Hicklin et al. 2023b DNAmix 2021 ICSA	Varied, including at least 13 STRmix versions	EPG profiles (.hid files)	52 (106)	8	1-2p 3-3p 3-4p 1-5p	88 to 486	≈2:1 ≈1:1:1 to 15:2:1 ≈8:6:4:1 to 30:17:12:1 ≈2:2:2:1:1
2023	Boodoosingh et al. (2024)	STRmix (v2.4.06, 2.5.11, 2.6.0, 2.6.3, 2.7.0) v2.9.1	EPG profiles (.hid files) generated in-house by 8 participating laboratories	8	20	2p to 4p	25 to 2100	N.E.S.

§ in the Barrio et al. 2018 study, a second sample with two males mixed 3:1 was also provided with Y-chromosome data

This information is intended to provide a snapshot of the factors explored in these studies. In situations where more datasets were received than laboratories participating, intralaboratory variation may have been considered as part of the study.

7.1. **Comments on Interlaboratory Studies**

Several trends can be extracted from Table S2.12.

First, most early studies focused on two-person mixtures, although a single three-person mixture was examined in each of the 1997, 1999, and 2001 NIST Mixed Stain Study series as well as the 2011 and 2012 GHEP studies. Only in the last decade or so has performance with low-level, high-contributor mixtures been studied, in large measure due to PGS use expanding across the community.

Second, a few of the earlier studies provided samples to participants to explore both measurement and interpretation aspects of the process either as stains (e.g., Duewer et al. 2001, Toscanini et al. 2016) or DNA extracts (e.g., Kline et al. 2003) whereas more recent studies have provided only DNA profile EPGs to examine interpretation variability across participants (e.g., Crespillo et al. 2014) or used publicly accessible EPGs (e.g., Bright et al. 2019a). Providing EPGs enables greater participation without having to prepare and disseminate physical samples but means that laboratory-specific measurement variability (e.g., DNA extraction and PCR amplification efficiency) cannot be assessed as part of the study. Also, when EPGs are provided, the interlaboratory study organizer has the burden of creating EPGs with multiple STR kits and trying to get consistent mixture ratios across the EPGs produced to avoid issues from sample differences (e.g., see Figure 1 in Butler et al. 2018). Since there are a wide variety of kits and protocols utilized across potential participants (e.g., Brinkac et al. 2023), study organizers need to make choices on what EPGs are provided. In addition, many analysts are uncomfortable analyzing data that was not collected under their laboratory protocols and choose not to participate because of this (Butler et al. 2018).

Third, PGS systems have played important roles in many of these interlaboratory studies. Some studies have focused on results when all participants use a specific pre-selected PGS system, such as LRmix (Prieto et al. 2014) or STRmix (Cooper et al. 2015, Bright et al. 2019a). Other studies have simply described what PGS system participants used (e.g., Mallinder et al. 2022, Hicklin et al. 2023a).

Fourth, many of the studies do not explicitly state the total DNA amount in the mixtures provided. Study organizers may focus on providing participants a consistent sample or set of EPG files and not describe every detail of their sample preparation in the subsequent publication. As with the other sources of information discussed in this supplemental document, future studies would benefit from having standard formats for data (see Key Takeaway #4.7 in NISTIR 8351).

7.1.1. **Some Overall Findings**

Initial NIST interlaboratory studies with the forensic DNA community, most of which are listed in Table S2.12 and conducted over 25 years ago, demonstrated: (1) that laboratories use instruments with different sensitivities, (2) experience and training play a part in effective mixture interpretation, and (3) the amount of input DNA affects the ability to detect the minor component in a mixture (Kline et al. 1997, Duewer et al. 2001, Kline et al. 2003, Duewer et al.

2004). For example, the publication for the 2001 NIST Mixed Stain Study #3 (MSS3) provided the following conclusion:

> "The MSS3 results also suggest that there are 10-fold differences in amplification, separation, and detection efficiencies among similar STR multiplex systems. Measurement particulars for a given laboratory at a given time period for particular instrumentation do not adequately predict the performance of nominally identical systems. This among-participant variability cannot be attributed to genetic methods or protocols, but rather, is associated with specific instruments, reagents, and analysts. This implies that STR multiplex DNA typing protocol and signal quality criteria should be performance-based and not prescriptive" (Kline et al. 2003).

Laboratories participating in the 2001 MSS3 study used the technology of their day – the ABI 310 single capillary or gel electrophoresis systems. Today, separation and detection systems are more sensitive with multiple possible parameters to be set by the analyst (see Brinkac et al. 2023), which has the potential to spread interlaboratory participants' protocol variation even more (see Hicklin et al. 2023a). Thus, many interlaboratory studies now emphasize the interpretation part of the process through providing data files rather than biological samples.

These findings influenced the design of later studies by NIST and others, and information from these studies have impacted the forensic DNA community in various ways. For example, a 2018 NIST publication noted:

> "Findings from the [NIST] MIX05 study influenced development of the Scientific Working Group on DNA Analysis Methods (SWGDAM) "SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Laboratories" released in 2010 (SWGDAM 2010) and updated in 2017 (SWGDAM 2017). Findings from the MIX13 study were initially shared at a DNA Technical Leader's Summit in November 2013 and have influenced the U.S. forensic community in recent years to move towards probabilistic genotyping approaches for complex DNA mixtures (see Bright et al. 2017). Findings from both studies have brought awareness of differences in approaches to DNA mixture interpretation and have highlighted the need for improved training and validation, which have hopefully led to improved protocols over the years" (Butler et al. 2018).

Comments and conclusions from studies in Table S2.12 are included below as well as subsequent sections that provide additional information on a few of the larger studies. Interested readers are invited to explore more details in the cited publications. Study volunteers are crucial to interlaboratory studies as acknowledged in NIST Mixed Stain Study #1:

> "Interlaboratory studies are made possible by the cooperation of many analysts and laboratory supervisors. We thank them for sharing with us their time and resources, and for their willingness to tackle our somewhat contrived samples. *Their voluntary and open participation in challenge exercises speaks as much to pursuit of analytical truth as to confidence in analytical systems*" (Duewer et al. 2001, emphasis added).

In the early 2010s, the European Forensic Genetics Network of Excellence (EuroForGen-NoE) received funding to advance DNA mixture interpretation and to provide training to forensic genetics experts in Europe. LRmix (a discrete PGS system) and EuroForMix (a continuous PGS

system) were developed from this effort (Gill & Haned 2013, Bleka et al. 2016a). Shortly after LRmix was deployed, training courses were held and an interlaboratory study conducted based on the training (Prieto et al. 2014).

This EuroForGen-NoE mixture study involved two exercises with 18 participating laboratories who were provided with two-person mixture data in conjunction with case scenarios, victim and suspect profiles, allele frequencies, probabilities of allele drop-out and drop-in, and LRmix software. The mixture in the first exercise had a single allele drop-out relative to the questioned contributor, the suspect, while the second exercise had 10 drop-out events relative to the questioned contributor, the victim. The study authors concluded "that the standardization of the probabilistic evaluation is possible, provided the same sets of hypotheses are compared, when suitable tools and training is provided to the DNA forensic experts" (Prieto et al. 2014).

The UK Forensic Science Regulator organized a 2014 study with eight participating laboratories that examined several two-person and three-person mixtures (Barber et al. 2015). This study found "a pressing need for a more consistent approach to mixture analysis, interpretation, and reporting, which impact on training, proficiency, and standards," which influenced the development of a 2017 judicial primer (Royal Society 2017) and a 2018 UK Forensic Science Regulator document on DNA mixture interpretation (UKFSR 2018). A follow-up study in 2018 demonstrated improved mixture interpretation across the same laboratories (Mallinder et al. 2022).

In 2014 and 2015, the Defense Forensic Science Center (DFSC) provided six DNA mixtures of varying difficulty to over 180 examiners from 55 participating forensic laboratories and found "significant intra- and inter-laboratory variation" (Aranda et al. 2015, Rogers et al. 2022). This variation was characterized via a novel allelic match score and a genotype interpretation metric. Based on results obtained, which preceded use of PGS systems, the study's authors concluded:

> "Two-person mixtures with signal peaks above stochastic threshold are generally interpretable, while three-person mixtures are currently beyond the scope or protocol limits for most participating examiners…There are, however, laboratories and participants that were able to interpret the difficult three-person mixtures and resolve genotypes for each contributor, even under very challenging conditions with nearly equivalent contributor ratios" (Rogers et al. 2022, p. 17).

In 2016, the Netherlands Forensic Institute (NFI) prepared 10 mixtures, most with replicate PCR amplifications, and sought input from 26 reporting officers across three European forensic laboratories to assess the level of inconsistency both within and between laboratories (Benschop et al. 2017a). This NFI-led study found "(almost) all participants from all three laboratories regarded a sample suitable for comparison to a reference DNA profile when the major contributor had no drop-out in any of the replicates, except when five-person mixtures were assessed..." In addition, "comparisons between the mixed profiles and reference profiles did not result in false inclusions…" The study authors reported:

> "Overall, variation was smaller within laboratories than between laboratories…*Similarity in answers depended mostly on the complexity of the profile sets*. For two profile sets all

> 26 [reporting officers] came to the same conclusion…For the other profile sets, laboratories use different criteria (complexity thresholds) for proceeding to profile comparison and/or [weight of evidence] assessment and these resulted in differences between the three laboratories" (Benschop et al. 2017a, emphasis added).

The NFI-led study shared the value and importance of these types of interlaboratory studies:

> "Studies like this help improve laboratory guidelines and explain differences between laboratories in court. Major differences within a laboratory should not exist if guidelines are adequately stringent, staff well-trained and thoroughly assessed prior to authorization, and if there is ongoing assessment of individual's output. Studies like this reveal differences among staff and allow root-cause analyses to identify core reasons for the differences which can then be addressed to ensure greater uniformity in service delivery. Such studies reveal differences between laboratories exposing the impact of differences in internal guidelines and training standards. Exposing these can assist in identifying improvement opportunities within laboratories to work toward best practice and improved service delivery to stakeholders. This is an ongoing process, and we encourage laboratories to regularly participate in, or organize such exercises using samples of known composition that allow for measuring exact performance" (Benschop et al. 2017a).

7.1.2. **GHEP-ISFG Collaborative Exercises**

The ISFG Spanish and Portuguese Speaking Working Group (GHEP) has held regular collaborative exercises involving DNA mixtures since 2004 (Garcia-Hirschfeld et al. 2006), and results from many of these studies have been published (e.g., Crespillo et al. 2014, Toscanini et al. 2016, Barrio et al. 2018). A Mixture Commission of the GHEP-ISFG was created in 2009 to administer mixture collaborative exercises "to contrast with each other their systematic analysis and interpretation of mixture profiles, as well as to check the statistical treatment used" and to have an educational aspect "to reveal some limiting factors[24] in the interpretation" (Crespillo et al. 2014).

Some observations from the 2010, 2011, and 2012 GHEP interlaboratory studies were that a majority of the errors are in the stutter position, that duplicate analysis using different STR kits helped resolve errors, that background case information was critical to select an appropriate set of propositions to assign LR values, and that allowing participants to set their own propositions led to significant differences in the LR assignments compared to using a common set of propositions across all participants (Crespillo et al. 2014).

The 2014 GHEP mixture study found that almost 20% of the inconsistencies observed came from insufficient electrophoretic resolution to resolve mixture components containing D12S391 "17.3" and "18" alleles that differed in length by a single nucleotide (Toscanini et al. 2016).

[24] The GHEP-ISFG authors state: "It is well known that behavior of mixture profiles at different phases of analysis (amplification, electrophoresis, results generation, interpretation and editing) is completely different from that of single profiles." (Crespillo et al. 2014, p. 71).

The 2015 GHEP collaborative exercise examined a 7:3:1 mixture provided as an EPG data file pre-analyzed with an analytical threshold of 50 RFU (Barrio et al. 2018). Many of the participating laboratories used LRmix Studio, an open-source discrete PGS system that considers alleles without peak height information along with allele drop-in and drop-out probabilities. Across 15 laboratories who may have used different assumptions in their LRmix Studio analysis, the assigned LR values differed by 12 orders of magnitude (see Table 1 in Barrio et al. 2018).

7.1.3. **STRmix Interlaboratory Studies**

The developers of STRmix have published two interlaboratory studies (Cooper et al. 2015, Bright et al. 2019a) and contributed to another (Riman et al. 2024a).

The 2014 study involved 20 participants from 12 different organizations assessing three Identifiler DNA profiles from casework (where the true number of contributors was unknown) that were provided as EPG files after analysis with a 50 RFU analytical threshold.

- For Case 1, all participants had a consistent assigned number of contributors (NoC=2), consistent propositions were used, and assigned LR point estimates varied by less than an order of magnitude (7.07×10^{10} to 8.49×10^{10}).
- For Case 2, the estimated number of contributors varied (eleven NoC=2, seven NoC=3, and two "unable to determine" or "inconclusive"), the propositions varied with the assigned NoC, and the overall assigned LR point estimates varied from 10^{7} to 10^{14}. Subdividing results by NoC increased the consistency somewhat depending on the assigned NoC with NoC=2 assigned LR values ranging from 1.65×10^{14} to 2.09×10^{14} and NoC=3 assigned LR values ranging from 7.43×10^{7} to 1.12×10^{10}.
- For Case 3, most participants (18 of 20; 90%) proposed NoC=3 and used consistent propositions (POI+U+U/U+U+U), which then produced assigned LR values mostly spanning two-orders of magnitude, ranging from 4.04×10^{10} to 4.37×10^{12} with one outlier (1.95×10^{8}). Variation in these three cases was later plotted in the second STRmix interlaboratory study (see Figure 2 in Bright et al. 2019a).

The authors of this interlaboratory study concluded:

> "This study demonstrates that for mixed DNA profiling results where the number of contributors is not ambiguous it is possible to achieve a standardized, consistent approach to the interpretation and statistical assessment of DNA evidence…[and]…the confident estimation and assignment of the number of contributors to DNA profiling results are essential to our ability to effectively interpret DNA profiles…" (Cooper et al. 2015) [25].

The 2018 study involved 174 participants from 42 laboratories assessing two complex mixtures using publicly accessible EPG files (Bright et al. 2019a). This study was designed in part as a

[25] The first STRmix interlaboratory study concluded: "[A] greatly improved degree of standardization can be achieved by implementation of the same probabilistic software within and between laboratories" (Cooper et al. 2015).

response to the variation reported with the GHEP-ISFG study (Barrio et al. 2018) discussed earlier and therefore many of the key known variables in the PGS process were held constant.[26]

Participants were invited to analyze the EPG files using pre-defined analytical thresholds, to label and model allelic and stutter peaks, to review provided case circumstances, to assign NoC to each sample, to develop suitable propositions, and to assign LR values in each case. About two-thirds of the participants used STRmix v2.4 and the remainder used STRmix v2.5. Assigned LR values were provided with the publication's supplemental materials (Bright et al. 2019a).

For Sample 1, which was a four-person mixture, five participants provided an inconclusive decision, nine reported an LR of 0 after assigning NoC=3, and the remainder assigned LR values ranging from 2.02×10^4 to 7.92×10^6. For Sample 2, which was a three-person mixture with higher DNA quantities, assigned LRs ranged from 2.21×10^{28} to 2.43×10^{29}. In terms of intra-laboratory variation, the study found that "the largest single laboratory (intra-laboratory) range of log(LR)s for Sample 1 was 2.09, which appears to be due to the use of different CE analysis methods by analysts."

The authors of this interlaboratory study concluded that this study "demonstrates a high level of repeatability and reproducibility among the participants. For those results that differed from the mode, the differences in LR were almost always minor or conservative" (Bright et al. 2019a).

In a 2024 publication, one of the developers of STRmix joined with the FBI Laboratory and NIST researchers to study the precision of the Markov chain Monte Carlo (MCMC) algorithms used for DNA profile interpretation (Riman et al. 2024a). This work analyzed 265 STR profiles (19 single-source, 59 two-person, 57 three-person, 55 four-person, 65 five-person, and 10 six-person mixtures) to study reproducibility across three laboratories running the same PGS version (STRmix v2.7) with identical input files, NoC settings, propositions, database of true and false donors, laboratory-specific STRmix settings, etc., but using a different computer and random number to start the MCMC simulations. Supplemental tables provide the assigned LRs from each laboratory and computed interlaboratory differences. Over 92% of assigned LRs fell within the same order of magnitude for the same input file, i.e., had a delta $\log_{10}$(LR) between 0 and 1, across the three laboratories (see Figure 4 in Riman et al. 2024a). Five key reasons were identified for poor precision: (1) non-convergence that can be diagnosed using the Gelman-Rubin statistic (Gelman & Rubin 1992, Russell et al. 2019), (2) saturated profiles, (3) profiles containing mixtures of equal contributions, (4) DNA profile quantity and/or quality, and (5) under assignment of the number of contributors. According to the authors, their main objective "was to highlight the types of profiles where the LR values were variable between repeat interpretations and discuss the causes of this variability" (Riman et al. 2024a).

[26] The second STRmix interlaboratory study noted: "…subjective decisions prior to application of the software can lead to a wide range in the reported LRs…In the study herein, we aimed to refine the sources of variation in the reported LR. In order to facilitate this study, the key known variables were set such as the allele frequency database, values for theta, and the various STRmix parameters controlling the biological modelling of peaks that in normal casework were defined by internal validation studies. Propositions were set by the participants based on the same case information" (Bright et al. 2019a).

7.1.4. **NIST MIX05 and MIX13**

The goal of the MIX05 and MIX13 studies performed by NIST was to examine sources of variability in interpretation rather than instrument sensitivity or amount of DNA being examined. Therefore, these studies involved sharing electronic files of DNA profiles with study participants rather than sharing biological samples. The STR profiles used for these two studies are still available on the NIST STRBase website at https://strbase.nist.gov/Information/Mixture_Studies. These profiles have been downloaded and used over the years for training purposes by many laboratories.

With MIX05, 69 laboratories interpreted two-person DNA mixtures from four mock sexual assault cases with different contributor ratios. The female victim DNA profile was supplied for each case, and participants were invited to deduce the male perpetrator DNA profile. No suspect(s) profiles or case scenarios were provided. Participants could use EPGs prepared from six STR kits (Identifiler, Profiler Plus, COfiler, SGM Plus, PowerPlex 16, PowerPlex 16 BIO).

- Case 1: male profile was the minor contributor (3:1 mixture)
- Case 2: male profile was the major contributor (3:1 mixture)
- Case 3: balanced mixture (1:1) where male profile lacked part of the sex-typing marker
- Case 4: male profile was a trace contributor (7:1 mixture) with a TPOX tri-allelic pattern

With MIX05 Case 1 where participants had to deduce the *minor component* genotypes with a substantial amount[27] of allele sharing, seven laboratories using the same STR kits reported results spanning 10 orders of magnitude (10^5 to 10^{15}). These laboratories used the 13 STRs present in Profiler Plus and COfiler and the same allele frequencies but with various detection thresholds (ranging from 75 to 150 RFUs) and different statistical approaches (random match probabilities from deduced minor contributor profile or combined probability of inclusion) (see Table 3 in Butler et al. 2018).

For MIX05 Case 2 where participants had to deduce the *major component* genotypes with very little[28] allele sharing, these same seven laboratories reported more consistent results (six were 10^{20} and one 10^{22}).

The publication on MIX05 concluded:

> "Thus, variation observed differs depending on the type and complexity of the mixture being evaluated as well as whether the reference profile being compared is a major or minor contributor to the mixture…MIX05 participants were highly accurate in deducing genotypes and fairly consistent in reporting statistics for a major contributor in a two-person mixture when there was very little overlap in genotypes present from the other contributor (Case 2). Accurately deducing minor contributor genotypes appeared to be

[27] MIX05 Case 1: Based on 15 STR loci with the Identifiler kit, there were 26 unshared alleles from a total of 39 (two loci exhibited only one allele, six loci exhibited two alleles, five loci exhibited three alleles, and two loci exhibited four alleles).
[28] MIX05 Case 2: Based on 15 STR loci with the Identifiler kit, there were 52 unshared alleles from a total of 55 (no loci exhibited only one allele, one locus exhibited two alleles, four loci exhibited three alleles, and ten loci exhibited four alleles).

more challenging and led to a larger spread in reported statistical values (e.g., Table 3) for the other cases examined (Case 1, Case 3, and Case 4)" (Butler et al. 2018).

In MIX13, 108 laboratories evaluated five case scenarios involving two, three, or four contributors, with some of the contributors potentially related. Suspect(s) profiles and case scenarios were provided. Participants could use EPGs prepared from two STR kits (Identifiler Plus or PowerPlex 16 HS).

- Case 1: two-person (1:1) mixture with a moderate degree of allele overlap to assess whether genotypes would be deduced, and which statistical approach would be taken
- Case 2: three-person (6:1.5:1) mixture with low template contributors to assess how potential allele dropout would be handled
- Case 3: three-person (7:2:1) mixture with a potential brother involved and where potential allele dropout was a possibility to assess impact of a related non-contributor
- Case 4: two-person (3.5:1) mixture with the person of interest as the minor component to explore separation of major and minor contributors
- Case 5: four-person (1:1:1:1) complex mixture that was designed with a high degree of allele sharing to assess how an uncertain number of contributors based on the number of observed alleles would be handled along with potential inclusion of an artificial suspect reference that was not part of the mixture to demonstrate problems with inappropriate application of a common combined probability of inclusion (CPI) interpretation approach

Discussion of MIX13 results often focuses on the contrived nature of Case 5 and its "over engineered" aspects (e.g., Buckleton et al. 2018) or the fact that 68% of the 108 participating laboratories incorrectly included the innocent "suspect C" with the CPI approaches used (e.g., Hampikian 2019). Despite the fact that significant variation in approaches and outcomes were observed from MIX13 participants, it would be inappropriate to make sweeping generalizations from this or any other interlaboratory collaborative exercise. Observations and limitations need to be considered in the context of the specific mixture examined given the experimental design and participants involved.

To illustrate the type of variation observed among the 108 MIX13 study participants at the time it was conducted in 2013, observations from the first case are included here. MIX13 Case 1 consisted of a two-person mixture designed to have equal amounts of each contributor (i.e., a 1:1 mixture ratio). All participants correctly included the reference profile "1A" and provided a statistic. Most of the laboratories inferred the genotype of the unknown contributor, often through conditioning on the victim given the case scenario. They provided a modified random match probability or used LR or CPI statistics. Across the 108 laboratory responses examining the same data, reported statistics spanned more than 20 orders of magnitude (Figure S2.1a) despite the fact that these mixtures were not complicated and did not possess any allele dropout (Figure S2.1b). Despite using the same EPG data, participants were making many different decisions that led to the variation observed.

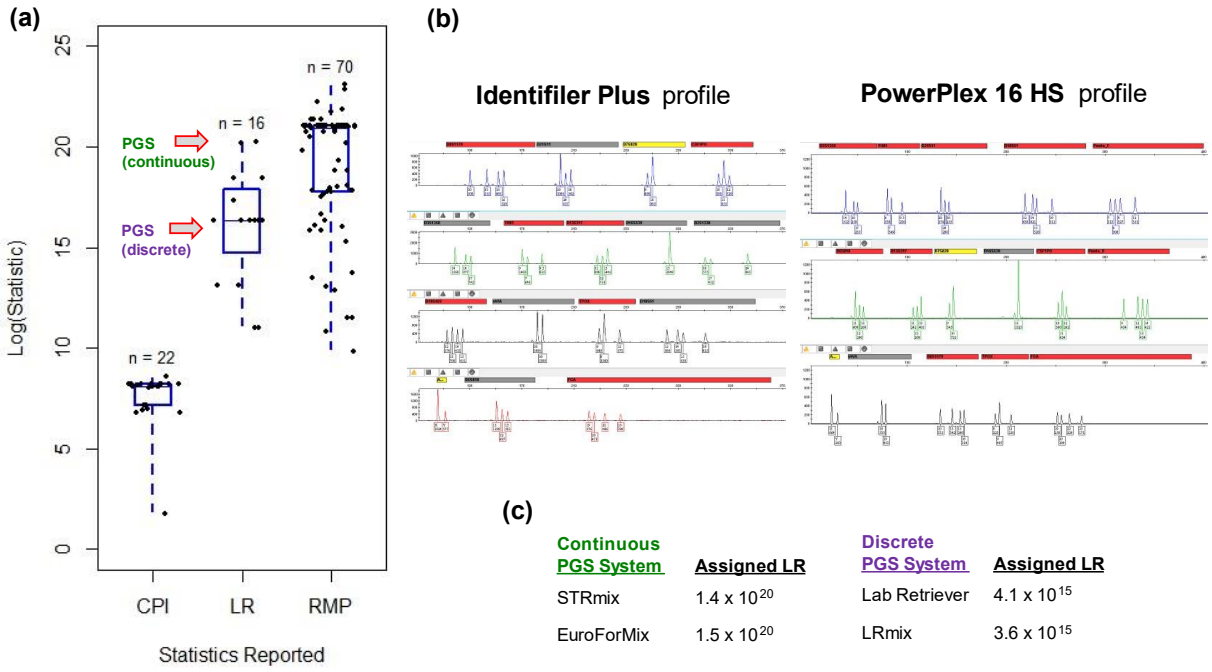


Figure S2.1. **(a)** Variation in 108 reported results for MIX13 Case 1, two-person (1:1) mixture, across statistical approaches of combined probability of inclusion (CPI), likelihood ratio (LR), and random match probability (RMP) using U.S. Caucasian allele frequencies. The vertical axis is in powers of 10 to reflect orders of magnitude as a $\log_{10}$(statistic). **(b)** Electropherogram (EPG) DNA profiles from Identifiler Plus and PowerPlex 16 HS evaluated by participants. **(c)** Results from continuous and discrete PGS systems later run in a single laboratory (Buckleton et al. 2018). Adapted from Butler et al. 2018, Figure 2 and Figure S2.

Given that many laboratories have moved to PGS in the past decade, the question could be asked what if a probabilistic genotyping approach had been used on this same sample? Four PGS systems were run *in a single laboratory* on the NIST MIX13 profiles to answer that question (Buckleton et al. 2018). For the MIX13 Case 1 mixture, two PGS systems using discrete models (Lab Retriever and LRmix) assigned LR values of 10^{15} while two PGS systems using continuous models (STRmix and EuroForMix) assigned LR values of around 10^{20} (Figure S2.1c). Clearly PGS approaches performed in a single laboratory make better use of the available data than manual approaches used in 2013. However, an interlaboratory study involving PGS systems (e.g., Bright et al. 2019a) is needed to assess community variation for DNA mixture interpretation using this approach. As described in the next section, this occurred with the DNAmix 2021 study a few years later.

7.1.5. **DNAmix 2021**

Since DNAmix 2021 was conducted fairly recently and was probably the largest and most comprehensive interlaboratory study to-date involving complex DNA mixtures, it will be discussed in greater detail and depth. This study was conducted in four phases from June 2021 through August 2022 by Noblis, Inc. (Reston, VA) and Bode Technology (Lorton, VA) with

financial support from a National Institute of Justice grant. Two articles describing the DNAmix 2021 findings have been published (Brinkac et al. 2023, Hicklin et al. 2023a) and a third one (Hicklin et al. 2023b) was in process of being published when this summary was finalized. According to the organizers, this study:

> "serves as the first large-scale study evaluating the extent of variation in interpretation and statistical analysis of DNA mixtures, specifically to: include results from current state-of-the-practice probabilistic genotyping software (PGS), with samples selected to be representative of the range of attributes found in actual DNA casework, using only real human DNA samples, and not restricted to any specific product or statistical approach" (Hicklin et al. 2023a).

Following registration on a website, DNAmix 2021 participants contributed responses in some or all of four phases: (1) an online policies and procedures (P&P) questionnaire to assess laboratory-specific protocols and parameters, (2) an online casework scenario questionnaire (CSQ) to assess laboratory/analyst approaches regarding analysis options that may vary depending upon the case scenario and the nature of their mixture casework, (3) a subtest where participants were provided electropherogram data from 12 mixtures (from a set of 21 possible mixtures) to assess suitability and number of contributors (NoC), and (4) a subtest where participants were provided electropherogram data from 8 mixtures along with DNA profiles of potential contributors to assess interpretation, comparison, and statistical analysis (ICSA).

A DNAmix Working Group and DNAmix Advisory Group collaboratively assisted the Bode/Noblis team with their study design including the scope, questions, and potential multiple-choice response categories. Thus, every effort was made to design an effective intra- and interlaboratory study to assess performance on casework-like DNA mixtures. This study introduced important features including: (a) gathering informed consent of participants, (b) guaranteeing anonymous participation, (c) requesting only qualified DNA analysts who would treat the study with the same diligence as used in operational casework, (d) permitting use of laboratory policies and procedures (as well as *capturing differences in policies and procedures among participating laboratories*), and (e) following quality assurance procedures including technical review of results.

In terms of gaining a meaningful set of participants and responses that would be reflective of community performance, the DNAmix 2021 instructions explain:

> "Participation is open to all forensic laboratories that conduct DNA mixture interpretation as part of their standard operating procedures (SOPs)...Participation in this study requires the participants to agree to use the same diligence in performing these analyses as used in operational casework, and to use their laboratory's SOPs in performing these analyses and conducting any quality assurance procedures required…" (see supplemental file to Brinkac et al. 2023).

The instructions explain further:

> "For purposes of this study, participants are laboratories not individuals… [although] laboratories will be permitted to register more than one participant." In addition,

"analysts involved must be qualified by the laboratory for operational mixture casework (not trainees)." Furthermore, "technical reviews and quality assurance procedures as outlined in the laboratory's SOPs should also be conducted…" (see supplemental file to Brinkac et al. 2023).

In total, 179 participants from 87 different laboratories participated in at least one phase of DNAmix 2021 (Brinkac et al. 2023):

- Phase 1 (P&P): 178 responses were reported by 178 participants from 86 laboratories,
- Phase 2 (CSQ): 163 responses were reported by 163 participants from 83 laboratories,
- Phase 3 (NoC): 1507 responses were provided by 134 participants from 67 laboratories,
- Phase 4 (ICSA): 765 responses were provided by 106 participants from 52 laboratories.

For the NoC portion, participants were assigned 12 mixtures from a set of 21 possible mixtures. For the ICSA portion, all participants received the same 8 mixtures. Thus, overall, there were 2272 assessments on 29 DNA mixtures (Hicklin et al. 2023a). Electropherograms of these 29 mixtures, which were created with four different amplification and capillary electrophoresis settings, have been archived as PDF images[29]. For the study though, these mixtures were supplied to participants as electropherograms in an .HID file format so that individual analytical thresholds and stutter filters could be applied to the data based on laboratory SOPs. A portion of the ICSA mixtures did not have the provided person of interest (POI) reference profile present in the mixture "so that the participants could make no assumptions regarding the presence or absence of the POIs" (Hicklin et al. 2023a).

A valuable contribution from this study via the P&P and CSQ phases was documentation of variation in methods and practices across the participant laboratories, which came from 29 U.S. states and 7 non-U.S. countries. Some highlights include (see Brinkac et al. 2023):

i. About three-quarters of laboratories terminated analysis prior to amplification depending upon the total DNA quantity, some with a threshold at 10 pg and some at 50 pg (section 2.4.1)

ii. Just over half of participating laboratories used at least one type of enhanced amplification method, such as concentration of the sample or increased injection time (section 2.4.2)

iii. Across 10 different STR kits, there were 27 different combinations of PCR cycle numbers (Table 3) and just over one-quarter routinely used replicate amplifications and most of these used all replicates in interpretation (Table 4)

iv. Across four different capillary electrophoresis instruments, there were 14 different injection voltages (Table 5) and 44 different injection times (Table 6)

v. For DNA profile interpretation, analytical thresholds ranged from 40 to 200 relative fluorescence units (RFUs) and stochastic thresholds from 150 to 1250 RFUs (section 2.4.5)

[29] See https://osf.io/b3mzw/ (accessed October 23, 2024)

vi. Nearly all laboratories manually assessed the number of contributors, most by considering the maximum allele count per locus (section 2.4.6 and Table 9)

vii. Nearly 90% limited their interpretation/comparison of DNA mixtures based upon a maximum total NoC, most commonly NoC=4 (Table 12)

viii. Most laboratories (71 of 86; 83%) reported some kind of likelihood ratio (continuous, binary, and/or semi-continuous) and (21 of 86; 24%) used combined probability of inclusion (CPI) for at least some cases (section 2.4.11 and Table 16)

ix. Most laboratories (52 of 86; 60%) used the NIST population databases for STR allele frequencies and over three-quarters routinely computed statistics for at least three populations (section 2.4.12)

x. At the time of the study, approximately two-thirds (58 laboratories) were using PGS for operational casework with 52 employing STRmix across 13 different versions ranging from v2.3 to v2.9 (section 2.4.13 and Table 17)

xi. Of those reporting LRs, almost half (31 laboratories) did not use a verbal equivalent in reporting (section 2.4.15) and almost all (80 laboratories) stated that they followed the 2017 SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories (Table 23)

xii. For most laboratories, all case information was provided to the DNA analyst up front (68 of 83; 82%) while some used a case manager or other individual to triage or restrict access to case information (9 of 83; 11%) (Table 24)

xiii. Those 69 laboratories participating in the casework scenario questionnaire that report LRs shared that they "do not often get information directly from the prosecution or the defense to help them formulate their propositions" (section 2.5.1)

xiv. The majority of laboratories encountered samples with up to 5 or more unknown contributors (section 2.5.6 and Table 27)

xv. When the case scenario suggested that the victim and the person of interest are first degree biological relatives (Table 29) or the POI alleged that a first degree biological relative, such as a sibling, was an alternative contributor (Table 30), about half of the participating laboratories did nothing differently and effectively performed statistical calculations as if they were considering unrelated individuals (section 2.5.8)

So, what was the impact of this level of variation across the participating laboratories?

This was a key question of the DNAmix 2021 study according to the lead investigator[30], who shared in a 2022 presentation: "If two laboratories are given the same mixture (as an electropherogram) and the same person of interest, how consistent are the statistical

[30] From a presentation given by Austin Hicklin at the February 2022 NIJ Symposium "Inter-laboratory Variation in Interpretation of DNA Mixtures Study: ***DNAmix 2021*** Phase 1: Policies & Procedures Questionnaire *Preliminary Results"* available at https://web.archive.org/web/20220806020641/https://dnamix.edgeaws.noblis.org/DNAMix_PnP-prelimResults-AAFS-NIJ-2022-02-17.pdf (accessed October 23, 2024).

responses and categorical interpretations? And what factors explain any differences in responses?"

From 2272 assessments of 29 DNA mixtures that were provided as EPGs to 134 participants in 67 forensic laboratories, the authors of the DNAmix 2021 study concluded:

> "Policies and procedures related to suitability and NoC varied notably among labs…if two labs following their standard operating procedures (SOPs) were given the same mixture, they agreed on whether the mixture was suitable for comparison 66% of the time… For labs following their SOPs, 79% of assessments of NoC were correct. When two different labs provided NoC responses, 63% of the time both labs were correct, and 7% of the time both labs were incorrect. Incorrect NoC assessments have been shown to affect statistical analyses in some cases, but do not necessarily imply inaccurate interpretations or conclusions…" (Hicklin et al. 2023a).

Whether a mixture EPG is suitable for interpretation is a decision each analyst makes based on their laboratory policies and perhaps their own level of perceived risk when those policies do not have clearly defined thresholds. This decision may center around the measured DNA quantity and/or the estimated NoC (e.g., see bullet points i and vii on the previous page). When a mixture is deemed not suitable, then no further analysis is conducted. Suitability becomes an important metric because failure to have reproducibility in suitability assessments across analysts and laboratories "in an operational context would mean receiving interpretation/analysis from one lab, and nothing from another" (Hicklin et al. 2023a).

The DNAmix 2021 study found that analysts from the same laboratory using the same SOP agreed on their suitability for a given mixture around 86% of the time and therefore disagreed approximately 14% of the time while analysts from different U.S. laboratories using the same SOP would disagree approximately 31% of the time on whether a given mixture from this study was suitable for interpretation (see Figure 3 in Hicklin et al. 2023a). Suitability responses varied across mixture types, although almost all participants found the five- and six-person mixtures not suitable (see Figure 1 in Hicklin et al. 2023a).

Extensive variation was also observed in assigned LRs, although publicly accessible details will need to await publication (Hicklin et al. 2023b). From what has been reported in public presentations on DNAmix 2021, extensive variation with DNA mixture interpretation and assigned LR values continues among interlaboratory study participants in spite of the fact that that a majority of laboratories are now utilizing PGS systems. This would suggest that even having a (mostly) common PGS system does not reduce the variability significantly, much like what was seen with proficiency test data.

There are many parameters to consider with mixture interpretation, and different analysts may apply these parameters differently, such that even when using many of the same components, such as a common PGS system, significant differences may be seen in an assigned LR value and sometimes even in the reported result.

7.2. **Additional Thoughts**

Over time, the interlaboratory comparison studies outlined in Table S2.12 have raised awareness of issues seen in DNA mixture interpretation across the participants of each study. These types of studies however cannot inform interested parties about the performance of the entire community or reliability of DNA mixture interpretation in a specific case of interest.

There have been a few additional published studies that are not listed in Table S2.12 that have compared results across analysts or laboratories in a more informal manner. For example, intralaboratory variation among 17 DNA analysts in the same laboratory was reported for mixture interpretation based on the presence or absence of contextual information about a criminal case (Dror & Hampikian 2011). In addition, a Netherlands study examined 19 DNA expert reports from 13 forensic institutes across seven countries and found important differences in content and conclusions when the same formal request, case context, and DNA profiles were provided (de Keijser et al. 2016).

Discussion from a 2018 NIST publication is worth repeating here:

> "These interlaboratory studies were not intended as a proficiency test but rather are designed as a training tool and an opportunity to discover the general performance across the community with the mixture scenarios being explored…[and]…studies like MIX05 and MIX13 may not always provide a full window into day-to-day performance in forensic laboratories. Variation observed and mistakes made in interlaboratory performance does not necessarily equate to innocent people being in jail – or the improper application of mixture interpretation in a specific case. Despite requests that the provided data be treated as if they involved real cases, results reported may not always have been handled as such. Some participants shared that results were provided back to NIST without the typical technical review that would be present before a real case report is released. Other laboratories may have conducted more extensive review than normal in reporting their results… [In addition,]…the intra-laboratory results suggest that training consistency may be an issue in some situations as different analysts in the same laboratory using the same protocol provided different results…In contrast, one large laboratory showed a great deal of consistency in their results (Table S5-(d)). This laboratory proved that it is possible to achieve consistency within a laboratory through a commitment to training and technical leadership. We have heard that an important outcome of this collaborative exercise is that some laboratories participating in the MIX13 study have implemented a routine mixture challenge to their analysts to help achieve better consistency. A regular review of DNA mixture interpretation performance within and across laboratories is expected to highlight areas for potential improvement" (Butler et al. 2018).

7.2.1. **Impact of Variable Responses on Potential Users**

A Netherlands study published in 2016 examined expert reports from 19 DNA practitioners coming from 13 forensic institutes across seven countries and found important differences in content and conclusions when the same formal request, case context, and DNA profiles were

provided (de Keijser et al. 2016). This study attempted to answer these two questions: (1) What type and magnitude of differences do we find when forensic DNA experts across institutes and across jurisdictions are handed an identical forensic case to report on?, and (2) If differences are substantive, what does this mean for the interpretation and value of the evidence by jurists?

This study involved three types of evidence (skin cells collected from the victim's shirt, saliva from a cigarette butt, and cells collected from under the victim's fingernails) with several types of two-person mixtures (a full major profile and a full minor profile from the shirt, a roughly equal mixture from the cigarette butt, and a full major profile with a partial secondary profile containing low peak heights from the fingernail). The largest difference in reported findings came with the mixed DNA profile from the victim's fingernails in which a low-quantity DNA contributor with allele dropout was present. The authors state:

> "On the one extreme is the *exclusion* of the suspect as a contributor to the profile. On the other extreme we find the statement that a match has been found between the suspect's profile and the few minor peaks in the mixed profile, and that it is 209 million times more likely if the victim and the suspect are contributors than if the victim and an unknown person are the contributors. Many of the differences described above were also found between reports originating from the same country…" (de Keijser et al. 2016, emphasis in the original).

Based on the information received from the 19 participating DNA experts, three reports were selected to illustrate differences in findings reported with the three pieces of evidence (shirt, cigarette butt, and fingernails). These reports were then translated into Dutch for use in studying potential jurist responses to them.

Potential jurists' perspectives were obtained through providing these three reports with the three pieces of evidence to graduate students taking a course about criminal evidence and inviting them to rate each piece of evidence on an 11-point scale (ranging from -5 "extremely exculpating," through 0, to +5 "extremely incriminating"). The students were also asked to consider comprehensibility of each report on a 7-point scale (e.g., ranging from 1 "do not understand it at all" to 7 "understand it completely") as they looked at the conclusions, the language, the logical structure, and whether the conclusions followed from the findings provided.

For example, with the DNA mixture recovered from under the victim's fingernails, the core elements of conclusions in the three reports were (1) an assigned LR of 5.7 million conditioning on the victim (i.e., H_1: V+S, H_2: V+U), (2) for minor contribution, minor DNA components were insufficient for identification purposes and "no male DNA was detected", and (3) for minor contribution, the suspect "eliminated as a significant contributor" with a possibility that suspect "contributed at a very low level to the result" that was not suitable for statistical evaluation.

In 69 "potential jurist" responses, the report from the DNA expert with an LR of 5.7 million was "considered incriminating by jurists" (1.53±2.38 on the 11-point scale) whereas the findings from the other two reports were "considered quite exculpating" (-3.46±1.81 and -2.88±1.95 on the 11-point scale). The authors' conclude:

> "This is a dramatic finding against the backdrop of all three reports having been written by actual DNA experts on the basis of exactly the same case materials and DNA profiles…In summary, expert reports written on the basis of identical case materials differ substantially, and, as a result have quite different meaning for jurists, depending on who authored the report" (de Keijser et al. 2016).

The authors of this study make some other important points in their discussion that can also relate to interlaboratory studies[31] in general. They wrote:

> "…different choices for interpreting and reporting DNA analysis can be made without those choices necessarily having to be false. The extent to which these choices are, should, or can be contested is not the scope of our study and up to the forensic DNA community to discuss. But even if each of the choices made by our forensic participants were totally defensible and justified in themselves, the reality remains that these choices result in widely different reports that may have real and far reaching consequences in court…While harmonization within and between jurisdictions may (eventually) solve the issue of same cases being treated differently in court as a result of differential forensic reporting, it is not the solution *per se*. In fact, harmonization in the absence of scientific consensus simply *hides* the actual underlying differences that are prevalent in current forensic practice. Here lies a great task for the forensic community, i.e., before harmonization to thoroughly (further) map the differences and work towards scientific consensus" (de Keijser et al. 2016, emphasis in the original).

[31] For example, citing the de Keijser et al. 2016 study findings in terms of observed differences between reports issued by participating laboratories regarding "extensiveness of the reports, explanation of technical issues, the use of explanatory appendices, level of reporting, use of context information, and most markedly, the type and content of the conclusions", the authors of a 2015 mixture study state: "Participating laboratories belonging to GHEP-ISFG behaved similarly in this collaborative exercise" (Barrio et al. 2018, quote from p. 162).

8. Summary

This supplemental document contains a summary of publicly accessible information regarding DNA mixture interpretation from five types of sources: (1) published developmental validation studies from STR typing kits, (2) published PGS studies, (3) PGS internal validation summaries from forensic laboratories, (4) proficiency test results, and (5) published interlaboratory studies. Currently, publicly accessible data does not have the necessary detail (including metadata, protocols, conditions, etc.) to enable transparency for an external and independent assessment of the degree of reliability of DNA mixture interpretation practices, including the use of probabilistic genotyping software systems.

As seen by the information summarized in this document and in the main report (NISTIR 8351), there is a growing body of scientific literature on DNA mixture interpretation. However, supporting data provided in this literature is not always sufficiently detailed for an independent review of claims surrounding the reliability of DNA mixture interpretation conducted at the sub-source level in the hierarchy of propositions (see Table 2.5 in NISTIR 8351). Such data and details, if required as part of the journal publication acceptance process, will assist in independent review of published articles.

It is encouraging that "the #1 journal in forensic genetics" (Kayser et al. 2024) included in a 2023 editorial several statements about the importance of accessible data (Kayser et al. 2023):

- "*Forensic Science International: Genetics* fully supports open science. Authors are encouraged to explore the various options that are available to them to ensure that their research is fully transparent from inception to completion."
- "Effective peer review and publication is heavily dependent on open accessibility to data, methods, and software. This allows reviewers and other readers to independently verify results and perform further analy[s]es, fostering continued review and exploration of new research questions even after publication. With their submission, authors must ensure that sufficient information is available for independent verification and replication of their findings…"
- In a check list for editors and peer-reviewers to follow, #3 states: "Are the conclusions supported by a comprehensive set of open-access data and/or open-source software and/or other documentation to ensure that the work is accessible and useful to the broader community to be published in the Journal?"

It will be interesting to see what the impact of these editorial considerations for publication in *Forensic Science International: Genetics* will have on the future of data accessibility relating to publications on DNA mixture interpretation in this journal.

In spite of limited information that is publicly accessible, it is recognized that forensic DNA laboratories have performed internal validation studies on their overall DNA interpretation procedures including use of PGS systems and have made decisions about implementing specific protocols for DNA mixture interpretation. In some situations, validation data containing DNA profiles of these DNA samples cannot be shared to protect the privacy of those individuals, who may not have provided informed consent for sharing their DNA profile.

An examination of publicly accessible proficiency test results involving DNA mixture interpretation found a very high success rate for correctly including or excluding associated reference samples when simple two-person mixtures when high quantities of DNA are used. With the few available three-person and four-person mixture assessments, particularly when low quantities of contributor DNA led to allele dropout, there were more false negatives (i.e., failure to include a true contributor) and inconclusive decisions. An analysis of results from four probabilistic genotyping proficiency tests, available only since 2022, observed variation in assigned likelihood ratios of more than 20 orders of magnitude. This variation may arise from different propositions, STR kits, PGS models, population databases, or reporting policies (e.g., use of a reporting cap of one billion as suggested by EWG 2024, pp. 85-87). There have been no known false inclusions using PGS (i.e., including a non-contributor in a mixture) based on proficiency test results examined thus far (see Table S2.7).

Interlaboratory studies have demonstrated a wide variety of approaches used for DNA mixture interpretation over the past 25 years – and that there is room for improvement if the desired goal is better consistency in DNA mixture interpretation across the community.

References

[NISTIR 8351-draft PCs] Public comments received on NISTIR 8351-DRAFT DNA Mixture Interpretation: A NIST Scientific Foundation Review. Published December 3, 2021. https://www.nist.gov/system/files/documents/2021/12/03/Public-Comments-NISTIR8351draft_0.pdf (accessed August 23, 2024).

Adamowicz MS, Rambo TN, Clarke JL (2022) Internal validation of MaSTR™ probabilistic genotyping software for the interpretation of 2-5 person mixed DNA profiles. *Genes* 13(8): 1429. https://doi.org/10.3390/genes13081429

Albright TD & Scurich N (2024) A call for open science in forensics. *Proceedings of the National Academy of Sciences of the United States of America* 121(24): e2321809121. https://doi.org/10.1073/pnas.2321809121

Alfonse LE, Garrett AD, Lun DS, Duffy KR, Grgicak CM (2018) A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt. *Forensic Science International: Genetics* 32:62-70. https://doi.org/10.1016/j.fsigen.2017.10.006

Alladio E, Omedei M, Cisana S, D'Amico G, Caneparo D, Vincenti M, Garofano P (2018) DNA mixtures interpretation - A proof-of-concept multi-software comparison highlighting different probabilistic methods' performances on challenging samples. *Forensic Science International: Genetics* 37:143-150. https://doi.org/10.1016/j.fsigen.2018.08.002

Aranda R (2015) A Large-scale Study of DNA Mixture Interpretation: Inter- and Intra-Laboratory Variability. *Proceedings of the 2015 International Symposium on Forensic Science Error Management* (NIST Special Publication 1206) Editor: JM Butler. Abstract available at https://doi.org/10.6028/NIST.SP.1206 (see p. 60) and slides available at https://www.nist.gov/system/files/documents/2016/11/22/large_scale_study_of_dna_mixture_interpretation.aranda-iv.crim1_.pdf. (accessed July 5, 2024).

Balding DJ (2013) Evaluation of mixed-source, low-template DNA profiles in forensic science. *Proceedings of the National Academy of Sciences of the United States of America* 110(30):12241-12246. https://doi.org/10.1073/pnas.1219739110

Ballantyne J, Hanson EK, Perlin MW (2013) DNA mixture genotyping by probabilistic computer interpretation of binomially-sampled laser capture cell populations: combining quantitative data for greater identification information. *Science & Justice* 53:103-114. https://doi.org/10.1016/j.scijus.2012.04.004

Barber M, Evett I, Pope S, Sullivan K, Whitaker J (2015) HOS/13/037 Forensic Science Regulator Codes of Practice and Conduct – Proficiency testing guidance for DNA mixtures. Final report. January 2015. Principal Forensic Services Ltd.

Barrio PA, Crespillo M, Luque JA, Aler M, Baeza-Richer C, Baldassarri L, Carnevali E, Coufalova P, Flores I, García O, García MA, González R, Hernández A, Inglés V, Luque GM, Mosquera-Miguel A, Pedrosa S, Pontes ML, Porto MJ, Posada Y, Ramella MI, Ribeiro T, Riego E, Sala A, Saragoni VG, Serrano A, Vannelli S (2018) GHEP-ISFG collaborative exercise on mixture profiles (GHEP-MIX06). Reporting conclusions:

results and evaluation. *Forensic Science International: Genetics* 35: 156-163. https://doi.org/10.1016/j.fsigen.2018.05.005

Bauer DW, Butt N, Hornyak JM, Perlin MW (2020) Validating TrueAllele® interpretation of DNA mixtures containing up to ten unknown contributors. *Journal of Forensic Sciences* 65(2):380-398. https://doi.org/10.1111/1556-4029.14204

Benschop CC, Haned H, de Blaeij TJ, Meulenbroek AJ, Sijen T (2012) Assessment of mock cases involving complex low template DNA mixtures: A descriptive study. *Forensic Science International: Genetics* 6(6):697-707. https://doi.org/10.1016/j.fsigen.2012.04.007

Benschop CC & Sijen T (2014) LoCIM-tool: An expert's assistant for inferring the major contributor's alleles in mixed consensus DNA profiles. *Forensic Science International: Genetics* 11:154-165. https://doi.org/10.1016/j.fsigen.2014.03.012

Benschop CC, Haned H, Yoo SY, Sijen T (2015a) Evaluation of samples comprising minute amounts of DNA. *Science & Justice* 55(5): 316-322. https://doi.org/10.1016/j.scijus.2015.05.002

Benschop CC, Haned H, Jeurissen L, Gill PD, Sijen T (2015b) The effect of varying the number of contributors on likelihood ratios for complex DNA mixtures. *Forensic Science International: Genetics* 19:92-99. https://doi.org/10.1016/j.fsigen.2015.07.003

Benschop CC, Connolly E, Ansell R, Kokshoorn B (2017a) Results of an inter and intra laboratory exercise on the assessment of complex autosomal DNA profiles. *Science & Justice* 57(1): 21-27. https://doi.org/10.1016/j.scijus.2016.10.001

Benschop CCG, van de Merwe L, de Jong J, Vanvooren V, Kempenaers M, van der Beek CP, Barni F, Reyes EL, Moulin L, Pene L, Haned H, Sijen T (2017b) Validation of SmartRank: A likelihood ratio software for searching national DNA databases with complex DNA profiles. *Forensic Science International: Genetics* 29:145-153. https://doi.org/10.1016/j.fsigen.2017.04.008

Benschop CCG, Nijveld A, Duijs FE, Sijen T (2019a) An assessment of the performance of the probabilistic genotyping software EuroForMix: Trends in likelihood ratios and analysis of Type I & II errors. *Forensic Science International: Genetics* 42:31-38. https://doi.org/10.1016/j.fsigen.2019.06.005

Benschop CCG, Hoogenboom J, Hovers P, Slagter M, Kruise D, Parag R, Steensma K, Slooten K, Nagel JHA, Dieltjes P, van Marion V, van Paassen H, de Jong J, Creeten C, Sijen T, Kneppers ALJ (2019b) DNAxs/DNAStatistX: Development and validation of a software suite for the data management and probabilistic interpretation of DNA profiles. *Forensic Science International: Genetics* 42:81-89. https://doi.org/10.1016/j.fsigen.2019.06.015

Benschop CCG, Hoogenboom J, Bargeman F, Hovers P, Slagter M, van der Linden J, Parag R, Kruise D, Drobnic K, Klucevsek G, Parson W, Berger B, Laurent FX, Faivre M, Ulus A, Schneider P, Bogus M, Kneppers ALJ, Sijen T (2020) Multi-laboratory validation of DNAxs including the statistical library DNAStatistX. *Forensic Science International: Genetics* 49:102390. https://doi.org/10.1016/j.fsigen.2020.102390

Bille TW, Weitz SM, Coble MD, Buckleton J, Bright JA (2014) Comparison of the performance of different models for the interpretation of low level mixed DNA profiles. *Electrophoresis* 35:3125-3133. https://doi.org/10.1002/elps.201400110

Bille T, Weitz S, Buckleton JS, Bright JA (2019) Interpreting a major component from a mixed DNA profile with an unknown number of minor contributors. *Forensic Science International: Genetics* 40:150-159. https://doi.org/10.1016/j.fsigen.2019.02.017

Bille T, Coble MD, Kalafut T, Buckleton J (2022) Study of CTS DNA proficiency tests with regard to DNA Mixture Interpretation: A NIST Scientific Foundation Review. *Genes* 13: 2171. https://doi.org/10.3390/genes13112171

Bleka Ø, Storvik G, Gill P (2016a) EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts. *Forensic Science International: Genetics* 21:35-44. https://doi.org/10.1016/j.fsigen.2015.11.008

Bleka Ø, Benschop CC, Storvik G, Gill P (2016b) A comparative study of qualitative and quantitative models used to interpret complex STR DNA profiles. *Forensic Science International: Genetics* 25:85-96. https://doi.org/10.1016/j.fsigen.2016.07.016

Bleka Ø, Prieto L, Gill P (2019) CaseSolver: An investigative open source expert system based on EuroForMix. *Forensic Science International: Genetics* 41:83-92. https://doi.org/10.1016/j.fsigen.2019.03.024

Boodoosingh S, Kelly H, Curran JM, Kalafut T (2024) An inter-laboratory comparison of probabilistic genotyping parameters and evaluation of performance on DNA mixtures from different laboratories. *Forensic Science International: Genetics* 71:103046. https://doi.org/10.1016/j.fsigen.2024.103046

Bright JA, Taylor D, Curran J, Buckleton J (2014) Searching mixed DNA profiles directly against profile databases. *Forensic Science International: Genetics* 9:102-110. https://doi.org/10.1016/j.fsigen.2013.12.001

Bright JA, Evett IW, Taylor D, Curran JM, Buckleton J (2015) A series of recommended tests when validating probabilistic DNA profile interpretation software. *Forensic Science International: Genetics* 14:125-131. https://doi.org/10.1016/j.fsigen.2014.09.019

Bright JA, Taylor D, McGovern C, Cooper S, Russell L, Abarno D, Buckleton J (2016) Developmental validation of STRmix, expert software for the interpretation of forensic DNA profiles. *Forensic Science International: Genetics* 23:226-239. https://doi.org/10.1016/j.fsigen.2016.05.007

Bright JA, Taylor D, Gittelson S, Buckleton J (2017) The paradigm shift in DNA profile interpretation. *Forensic Science International: Genetics* 31:e24-e32. https://doi.org/10.1016/j.fsigen.2017.08.005

Bright JA, Richards R, Kruijver M, Kelly H, McGovern C, Magee A, McWhorter A, Ciecko A, Peck B, Baumgartner C, Buettner C, McWilliams S, McKenna C, Gallacher C, Mallinder B, Wright D, Johnson D, Catella D, Lien E, O'Connor C, Duncan G, Bundy J, Echard J, Lowe J, Stewart J, Corrado K, Gentile S, Kaplan M, Hassler M, McDonald N, Hulme P, Oefelein RH, Montpetit S, Strong M, Noel S, Malsom S,

Myers S, Welti S, Moretti T, McMahon T, Grill T, Kalafut T, Greer-Ritzheimer MM, Beamer V, Taylor DA, Buckleton JS (2018) Internal validation of STRmix™– a multi laboratory response to PCAST. *Forensic Science International: Genetics* 34:11-24. https://doi.org/10.1016/j.fsigen.2018.01.003

Bright JA, Cheng K, Kerr Z, McGovern C, Kelly H, Moretti TR, Smith MA, Bieber FR, Budowle B, Coble MD, Alghafri R, Allen PS, Barber A, Beamer V, Buettner C, Russell M, Gehrig C, Hicks T, Charak J, Cheong-Wing K, Ciecko A, Davis CT, Donley M, Pedersen N, Gartside B, Granger D, Greer-Ritzheimer M, Reisinger E, Kennedy J, Grammer E, Kaplan M, Hansen D, Larsen HJ, Laureano A, Li C, Lien E, Lindberg E, Kelly C, Mallinder B, Malsom S, Yacovone-Margetts A, McWhorter A, Prajapati SM, Powell T, Shutler G, Stevenson K, Stonehouse AR, Smith L, Murakami J, Halsing E, Wright D, Clark L, Taylor DA, Buckleton J (2019a) STRmix™ collaborative exercise on DNA mixture interpretation. *Forensic Science International: Genetics* 40:1-8. https://doi.org/10.1016/j.fsigen.2019.01.006

Bright JA, Taylor D, Kerr Z, Buckleton J, Kruijver M (2019b) The efficacy of DNA mixture to mixture matching. *Forensic Science International: Genetics* 41:64-71. https://doi.org/10.1016/j.fsigen.2019.02.020

Brinkac LM, Richetelli N, Davoren JM, Bever RA, Hicklin RA (2023) DNAmix 2021: Laboratory policies, procedures, and casework scenarios summary and dataset. *Data in Brief* 48:109150. https://doi.org/10.1016/j.dib.2023.109150

Buckleton J (2009) Validation issues around DNA typing of low level DNA. *Forensic Science International: Genetics* 3(4):255-260. https://doi.org/10.1016/j.fsigen.2009.04.003

Buckleton JS, Bright J-A, Cheng K, Budowle B, Coble MD (2018) NIST Interlaboratory studies involving DNA mixtures (MIX13): a modern analysis. *Forensic Science International: Genetics* 37:172-179. https://doi.org/10.1016/j.fsigen.2018.08.014

Buckleton JS, Pugh SN, Bright JA, Taylor DA, Curran JM, Kruijver M, Gill P, Budowle B, Cheng K (2020) Are low LRs reliable? *Forensic Science International: Genetics* 49:102350. https://doi.org/10.1016/j.fsigen.2020.102350

Buckleton J, Bright JA, Taylor D, Wivell R, Bleka Ø, Gill P, Benschop C, Budowle B, Coble M (2022) Re: Riman et al. Examining performance and likelihood ratios for two likelihood ratio systems using the PROVEDIt dataset. *Forensic Science International: Genetics* 59: 102709. https://doi.org/10.1016/j.fsigen.2022.102709

Buckleton J, Susik M, Curran JM, Cheng K, Taylor D, Bright JA, Kelly H, Wivell R (2024) A diagnosis of the primary difference between EuroForMix and STRmix™. *Journal of Forensic Sciences* 69(1):40–51. https://doi.org/10.1111/1556-4029.15387

Budowle B & Sajantila A (2023) Revisiting informed consent in forensic genomics in light of current technologies and the times. *International Journal of Legal Medicine* 137(2):551-565. https://doi.org/10.1007/s00414-023-02947-w

Butler JM, Buel E, Crivellente F, McCord BR (2004). Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis* 25:1397-1412. https://doi.org/10.1002/elps.200305822

Butler JM (2011) *Advanced Topics in Forensic DNA Typing: Methodology*. Elsevier Academic Press: San Diego.

Butler JM, Kline MC, Coble MD (2018) NIST interlaboratory studies involving DNA mixtures (MIX05 and MIX13): variation observed and lessons learned. *Forensic Science International: Genetics* 37: 81-94. https://doi.org/10.1016/j.fsigen.2018.07.024

Chapman W, Hicklin A, Taylor M (2023) Beginners Guide to Biometric and Forensic Science Human Subjects Research Protections. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) NIST SP 1289. https://doi.org/10.6028/NIST.SP.1289

Cheng K, Bleka Ø, Gill P, Curran J, Bright J-A, Taylor D, Buckleton J (2021) A comparison of likelihood ratios obtained from EuroForMix and STRmix. *Journal of Forensic Sciences* 66(6):2138-2155. https://doi.org/10.1111/1556-4029.14886

Chin JM & Ibaviosa CM (2022) Beyond CSI: Calibrating public beliefs about the reliability of forensic science through openness and transparency. *Science & Justice* 62(3):272-283. https://doi.org/10.1016/j.scijus.2022.02.006

Collins PJ, Hennessy LK, Leibelt CS, Roby RK, Reeder DJ, Foxall PA (2004) Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFlSTR Identifiler PCR Amplification Kit. *Journal of Forensic Sciences* 49(6):1265-1277.

Cooper S, McGovern C, Bright JA, Taylor D, Buckleton J (2015) Investigating a common approach to DNA profile interpretation using probabilistic software. *Forensic Science International: Genetics* 16:121-131. https://doi.org/10.1016/j.fsigen.2014.12.009

Crespillo M, Barrio PA, Luque JA, Alves C, Aler M, Alessandrini F, Andrade L, Barretto RM, Bofarull A, Costa S, García MA, García O, Gaviria A, Gladys A, Gorostiza A, Hernández A, Herrera M, Hombreiro L, Ibarra AA, Jiménez MJ, Luque GM, Madero P, Martínez-Jarreta B, Masciovecchio MV, Modesti NM, Moreno F, Pagano S, Pedrosa S, Plaza G, Prat E, Puente J, Rendo F, Ribeiro T, Sala A, Santamaría E, Saragoni VG, Whittle MR (2014) GHEP-ISFG collaborative exercise on mixture profiles of autosomal STRs (GHEP-MIX01, GHEP-MIX02 and GHEP-MIX03): results and evaluation. *Forensic Science International: Genetics* 10:64-72. https://doi.org/10.1016/j.fsigen.2014.01.009

[CTS 22-5882] Collaborative Testing Services, Inc. (2023) DNA Interpretation Test No. 22-5882 Summary Report. Available at https://cts-forensics.com/reports/22-5882_Web.pdf (accessed October 19, 2024).

[CTS 24-5901/2] Collaborative Testing Services, Inc. (2024) Probabilistic Genotyping Test No. 24-5901/2 Summary Report. Originally released on May 31, 2024, and revised on July 24, 2024, with an updated Summary Comments section. Available at https://cts-forensics.com/reports/24-5901.2_Web.pdf (accessed October 19, 2024).

de Keijser JW, Malsch M, Luining ET, Weulen Kranenbarg M, Lenssen DJHM (2016) Differential reporting of mixed DNA profiles and its impact on jurists' evaluation of evidence. An international analysis. *Forensic Science International: Genetics* 23:71-82. https://doi.org/10.1016/j.fsigen.2016.03.006

[DFS Zoom Study] District of Columbia Department of Forensic Sciences (2022) The Zoom Study: Additional Guidelines for Interpretation of Mixtures and Low Level Data Using GlobalFiler on the 3500/3500xL and/or STRmix 2.4. Available at https://dfs.dc.gov/sites/default/files/dc/sites/dfs/page_content/attachments/STRmix%20v2.4%20Supplemental%20Interpretation%20Study.pdf (accessed October 21, 2024).

Dror IE & Hampikian G (2011) Subjectivity and bias in forensic DNA mixture interpretation. *Science & Justice* 51(4):204-208. https://doi.org/10.1016/j.scijus.2011.08.004

Duewer DL, Kline MC, Redman JW, Newall PJ, Reeder DJ (2001) NIST Mixed Stain Studies #1 and #2: interlaboratory comparison of DNA quantification practice and short tandem repeat multiplex performance with multiple-source samples. *Journal of Forensic Sciences* 46:1199-1210.

Duewer DL, Kline MC, Redman JW, Butler JM (2004) NIST Mixed Stain Study #3: signal intensity balance in commercial short tandem repeat multiplexes. *Analytical Chemistry* 76:6928-6934. https://doi.org/10.1021/ac049178k

Duke KR & Myers SP (2020) Systematic evaluation of STRmix performance on degraded DNA profile data. *Forensic Science International: Genetics* 44:102174. https://doi.org/10.1016/j.fsigen.2019.102174

Duke K, Cuenca D, Myers S, Wallin J (2022) Compound and conditioned likelihood ratio behavior within a probabilistic genotyping context. *Genes* 13(11): 2031. https://doi.org/10.3390/genes13112031

Duke K, Myers S, Cuenca D, Wallin J (2023) Improving the utilization of STRmix™ variance parameters as semi-quantitative profile modeling metrics. *Genes* 14(1):102. https://doi.org/10.3390/genes14010102

Edge MD & Matthews JN (2022) Open practices in our science and our courtrooms. *Trends in Genetics* 38(2):113-115. https://doi.org/10.1016/j.tig.2021.09.010

Ensenberger MG, Thompson J, Hill B, Homick K, Kearney V, Mayntz-Press KA, Mazur P, McGuckian A, Myers J, Raley K, Raley SG, Rothove R, Wilson J, Wieczorek D, Fulmer PM, Storts DR, Krenke BE (2010) Developmental validation of the PowerPlex 16 HS System: an improved 16-locus fluorescent STR multiplex. *Forensic Science International: Genetics* 4(4):257-264. https://doi.org/10.1016/j.fsigen.2009.10.007

Ensenberger MG, Hill CR, McLaren RS, Sprecher CJ, Storts DR (2014) Developmental validation of the PowerPlex(®) 21 System. *Forensic Science International: Genetics* 9:169-178. https://doi.org/10.1016/j.fsigen.2013.12.005

Ensenberger MG, Lenz KA, Matthies LK, Hadinoto GM, Schienman JE, Przech AJ, Morganti MW, Renstrom DT, Baker VM, Gawrys KM, Hoogendoorn M, Steffen CR, Martín P, Alonso A, Olson HR, Sprecher CJ, Storts DR (2016) Developmental validation of the PowerPlex(®) Fusion 6C System. *Forensic Science International: Genetics* 21:134-144. https://doi.org/10.1016/j.fsigen.2015.12.011

[ESR 2021] ESR Response to NISTIR 8351-DRAFT DNA Mixture Interpretation: A NIST Scientific Foundation Review (dataset posted on July 30, 2021). Available at

https://research.esr.cri.nz/articles/dataset/ESR_response_to_NISTIR_8351_-_DRAFT_DNA_Mixture_Interpretation_A_NIST_Scientific_Foundation_Review/15062907 (accessed October 21, 2024).

[EWG 2024] Expert Working Group; Taylor M, Romsos E, Ballantyne K, Moore Boswell D, Busey T. (2024) *Forensic DNA Interpretation and Human Factors: Improving the Practice Through a Systems Approach*. NIST Interagency Report (NISTIR) 8503, National Institute of Standards and Technology, Gaithersburg, MD. Available at https://doi.org/10.6028/NIST.IR.8503, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=956787 (Accessed May 10, 2024)

[GAO 2021] United States Government Accountability Office (2021) Technology Assessment: Forensic Technology: Algorithms Strengthen Forensic Analysis, but Several Factors Can Affect Outcomes. GAO-21-435SP. Available at https://www.gao.gov/assets/gao-21-435sp.pdf (accessed August 27, 2024).

Garcia-Hirschfeld J, Alsonso A, Garcia O, Amorim A, Gomez J (2006) 2004-2005 GEP proficiency testing programs: Special emphasis on the interlaboratory analysis of mixed stains. *Progress in Forensic Genetics 11 - International Congress Series* 1288:855-857. https://doi.org/10.1016/j.ics.2005.09.167

Gelman A & Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statistical Science* 7(4):457-472.

Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, Morling N, Prinz M, Schneider PM, Weir BS (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Science International* 160:90-101. https://doi.org/10.1016/j.forsciint.2006.04.009

Gill P & Haned H (2013) A new methodological framework to interpret complex DNA profiles using likelihood ratios. *Forensic Science International: Genetics* 7(2):251-263. https://doi.org/10.1016/j.fsigen.2012.11.002

Green RL, Lagacé RE, Oldroyd NJ, Hennessy LK, Mulero JJ (2013) Developmental validation of the AmpFℓSTR® NGM SElect™ PCR Amplification Kit: A next-generation STR multiplex with the SE33 locus. *Forensic Science International: Genetics* 7(1):41-51. https://doi.org/10.1016/j.fsigen.2012.05.012

Greenspoon SA, Schiermeier-Wood L, Jenkins BC (2015) Establishing the limits of TrueAllele® Casework: a validation study. *Journal of Forensic Sciences* 60(5):1263-1276. https://doi.org/10.1111/1556-4029.12810

Greenspoon SA, Schiermeier-Wood L, Jenkins BC (2024) A tale of two PG systems: A comparison of the two most widely used continuous probabilistic genotyping systems in the United States. *Journal of Forensic Sciences* 69(5):1840-1860. https://doi.org/10.1111/1556-4029.15571

Hampikian G (2019) Correcting forensic DNA errors. *Forensic Science International: Genetics* 41:32-33. https://doi.org/10.1016/j.fsigen.2019.03.005

Haned H, Benschop CC, Gill PD, Sijen T (2015) Complex DNA mixture analysis in a forensic context: evaluating the probative value using a likelihood ratio model. *Forensic Science International: Genetics* 16:17-25. https://doi.org/10.1016/j.fsigen.2014.11.014

Haned H, Gill P, Lohmueller K, Inman K, Rudin N (2016) Validation of probabilistic genotyping software for use in forensic DNA casework: Definitions and illustrations. *Science & Justice* 56(2):104-108. https://doi.org/10.1016/j.scijus.2015.11.007

Hicklin RA, Richetelli N, Emerick BL, Bever RA, Davoren JM (2023a) Variation in assessments of suitability and number of contributors for DNA mixtures. *Forensic Science International: Genetics* 65: 102892. https://doi.org/10.1016/j.fsigen.2023.102892

Hicklin RA, Richetelli N, Bever RA, Davoren JM (2023b) Inter-laboratory variation in interpretations, comparisons, and statistical analyses of DNA mixtures. (manuscript in preparation)

Holland MM, Tiedge TM, Bender AJ, Gaston-Sanchez SA, McElhoe JA (2022) MaSTR: an effective probabilistic genotyping tool for interpretation of STR mixtures associated with differentially degraded DNA. *International Journal of Legal Medicine* 136(2): 433-446. https://doi.org/10.1007/s00414-021-02771-0

Huffman K, & Ballantyne J (2022) Probabilistic genotyping of single cell replicates from mixtures involving first-degree relatives prevents the false inclusions of non-donor relatives. *Genes* 13(9):1658. https://doi.org/10.3390/genes13091658

Huffman K & Ballantyne J (2023) Validation of probabilistic genotyping software for single cell STR analysis. Genes 14(3): 674. https://doi.org/10.3390/genes14030674

[Jefferson County] ESR (2019) Internal Validation of STRmix v2.6 for the analysis of GlobalFiler profiles at the Jefferson County Regional Crime Laboratory. Available at https://indefenseof.us/uploads/Jefferson-County-STRmix-Validation-V2.6-V2.6.3.pdf (accessed October 22, 2024).

Kalafut T, Schuerman C, Sutton J, Faris T, Armogida L, Bright JA, Buckleton J, Taylor D (2018) Implementation and validation of an improved allele specific stutter filtering method for electropherogram interpretation. *Forensic Science International: Genetics* 35:50-56. https://doi.org/10.1016/j.fsigen.2018.03.016

Kalafut T, Pugh S, Gill P, Abbas S, Semaan M, Mansour I, Curran J, Bright JA, Hicks T, Wivell R, Buckleton J (2022a) A mixed DNA profile controversy revisited. *Journal of Forensic Sciences* 67(1): 128-135. https://doi.org/10.1111/1556-4029.14912

Kalafut T, Bright JA, Taylor D, Buckleton J (2022b) Investigation into the effect of mixtures comprising related people on non-donor likelihood ratios, and potential practises to mitigate providing misleading opinions. *Forensic Science International: Genetics* 59:102691. https://doi.org/10.1016/j.fsigen.2022.102691

Kayser M, Gill P, Parson W, Gusmão L, Linacre A, Vallone P, Carracedo A (2023) Editorial considerations for publication in Forensic Science International: Genetics. *Forensic Science International: Genetics* 65:102877. https://doi.org/10.1016/j.fsigen.2023.102877

Kayser M, Gusmão L, Linacre A, Parson W, Vallone P, Carracedo A (2024) Forensic Science International: Genetics is the leading journal in the field of legal medicine. *Forensic Science International: Genetics* 72:103097. https://doi.org/10.1016/j.fsigen.2024.103097

Kelly H, Bright JA, Kruijver M, Cooper S, Taylor D, Duke K, Strong M, Beamer V, Buettner C, Buckleton J (2018) A sensitivity analysis to determine the robustness of STRmix™ with respect to laboratory calibration. *Forensic Science International: Genetics* 35:113-122. https://doi.org/10.1016/j.fsigen.2018.04.009

Kelly H, Bright JA, Kruijver M, Taylor D, Buckleton J (2022a) The effect of a user selected number of contributors within the LR assignment. *Australian Journal of Forensic Sciences* 54(4):450-463. https://doi.org/10.1080/00450618.2020.1865456

Kelly H, Coble M, Kruijver M, Wivell R, Bright JA (2022b) Exploring likelihood ratios assigned for siblings of the true mixture contributor as an alternate contributor. *Journal of Forensic Sciences* 67(3): 1167-1175. https://doi.org/10.1111/1556-4029.15020

Kline MC, Duewer DL, Newall P, Redman JW, Reeder DJ, Richard M (1997) Interlaboratory evaluation of short tandem repeat triplex CTT. *Journal of Forensic Sciences* 42(5):897-906.

Kline MC, Redman JW, Duewer DL, Reeder DJ (1999) Results from the 1999 NIST Mixed-Stain Study #2: DNA quantitation, differential extraction, and identification of the unknown contributors. *Proceedings of the International Symposium on Human Identification* (Promega Corporation). Available at https://www.promega.com/products/pm/genetic-identity/ishi-conference-proceedings/10th-ishioral-presentations/. (accessed August 23, 2024).

Kline MC, Duewer DL, Redman JW, Butler JM (2003) NIST mixed stain study 3: DNA quantitation accuracy and its influence on short tandem repeat multiplex signal intensity. *Analytical Chemistry* 75:2463-2469. https://doi.org/10.1021/ac026410i

Koehler JJ (2013) Proficiency tests to estimate error rates in the forensic sciences. *Law, Probability and Risk* 12(1):89-98. https://doi.org/10.1093/lpr/mgs013

Kraemer M, Prochnow A, Bussmann M, Scherer M, Peist R, Steffen C (2017) Developmental validation of QIAGEN Investigator® 24plex QS Kit and Investigator® 24plex GO! Kit: Two 6-dye multiplex assays for the extended CODIS core loci. *Forensic Science International: Genetics* 29:9-20. https://doi.org/10.1016/j.fsigen.2017.03.012

Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ, Tomsey CS, Zachetti JM, Masibay A, Rabbach DR, Amiott EA, Sprecher CJ (2002) Validation of a 16-locus fluorescent multiplex system. *Journal of Forensic Sciences* 47(4):773-785.

Lazaruk K, Walsh PS, Oaks F, Gilbert D, Rosenblum BB, Menchen S, Scheibler D, Wenz HM, Holt C, Wallin J (1998) Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. *Electrophoresis* 19(1):86-93. https://doi.org/10.1002/elps.1150190116

Lin M-H, Bright J-A, Pugh SN, Buckleton JS (2020) The interpretation of mixed DNA profiles from a mother, father, and child trio. *Forensic Science International: Genetics* 44:102175. https://doi.org/10.1016/j.fsigen.2019.102175

Lucassen A, Ehlers K, Grobler PJ, Brenner CH (2021) Evaluating Mixture Solution- rapid and non-MCMC probabilistic mixture analysis. *International Journal of Legal Medicine* 135(6): 2275-2284. https://doi.org/10.1007/s00414-021-02680-2

Ludeman MJ, Zhong C, Mulero JJ, Lagacé RE, Hennessy LK, Short ML, Wang DY (2018) Developmental validation of GlobalFiler™ PCR amplification kit: a 6-dye multiplex assay designed for amplification of casework samples. *International Journal of Legal Medicine* 132(6):1555-1573. https://doi.org/10.1007/s00414-018-1817-5

[LVMPD] ESR (2020) Internal Validation of STRmix v2.6 (QIAGEN Investigator 24plex QS with 3500xl) Las Vegas Metropolitan Police Department (32 pages). Available at https://indefenseof.us/uploads/LVMPD-Summary.pdf (accessed October 21, 2024); see also PC12 in 8351-draft PC

Mallinder B, Pope S, Thomson J, Beck LA, McDonald A, Ramsbottom D, Court DS, Vanhinsbergh D, Barber M, Evett I, Sullivan K, Whitaker J (2022) Interpretation and reporting of mixed DNA profiles by seven forensic laboratories in the UK and Ireland. *Forensic Science International: Genetics* 58: 102674. https://doi.org/10.1016/j.fsigen.2022.102674

Manabe S, Morimoto C, Hamano Y, Fujimoto S, Tamaki K (2017) Development and validation of open-source software for DNA mixture interpretation based on a quantitative continuous model. *PLoS ONE* 12(11):e0188183. https://doi.org/10.1371/journal.pone.0188183

Manabe S, Fukagawa T, Fujii K, Mizuno N, Sekiguchi K, Akane A, Tamaki K (2022) Development and validation of Kongoh ver. 3.0.1: Open-source software for DNA mixture interpretation in the GlobalFiler system based on a quantitative continuous model. *Legal Medicine* 54: 101972. https://doi.org/10.1016/j.legalmed.2021.101972

Marciano MA & Maynard HP (2023) Enhancing research and collaboration in forensic science: A primer on data sharing. *Forensic Science International: Synergy* 6:100323. https://doi.org/10.1016/j.fsisyn.2023.100323

Marciano MA, Hall A, Mozayani A, Cromp T, Maynard HP (2023) Enhancing research and collaboration in forensic science: A primer on human subjects' research protection. *Forensic Science International: Synergy* 7:100443. https://doi.org/10.1016/j.fsisyn.2023.100443

McCarthy-Allen M, Bleka Ø, Ypma R, Gill P, Benschop C (2024) 'Low' LRs obtained from DNA mixtures: On calibration and discrimination performance of probabilistic genotyping software. *Forensic Science International: Genetics* 73:103099. https://doi.org/10.1016/j.fsigen.2024.103099

McGovern C, Cheng K, Kelly H, Ciecko A, Taylor D, Buckleton JS, Bright JA (2020) Performance of a method for weighting a range in the number of contributors in probabilistic genotyping. *Forensic Science International: Genetics* 48:102352. https://doi.org/10.1016/j.fsigen.2020.102352

Mitchell AA, Tamariz J, O'Connell K, Ducasse N, Budimlija Z, Prinz M, Caragine T (2012) Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in. *Forensic Science International: Genetics* 6(6):749-761. https://doi.org/10.1016/j.fsigen.2012.08.007

Moretti TR, Just RS, Kehl SC, Willis LE, Buckleton JS, Bright JA, Taylor DA, Onorato AJ (2017) Internal validation of STRmix for the interpretation of single source and mixed DNA profiles. *Forensic Science International: Genetics* 29:126-144. https://doi.org/10.1016/j.fsigen.2017.04.004

Myers S (2021) Searching CODIS with binary conversions of STRmix interpretations. *Forensic Science International: Genetics* 55:102569. https://doi.org/10.1016/j.fsigen.2021.102569

[NISTIR 8225] Butler JM, Iyer H, Press R, Taylor MK, Vallone PM, Willis S (2020) NISTIR 8225: NIST Scientific Foundation Reviews. Available at https://nvlpubs.nist.gov/nistpubs/ir/2020/NIST.IR.8225.pdf (accessed July 5, 2024).

[NISTIR 8351] Butler JM, Iyer H, Press R, Taylor MK, Vallone PM, Willis S (2024) DNA Mixture Interpretation: A NIST Scientific Foundation Review. (National Institute of Standards and Technology, Gaithersburg, MD), NISTIR 8351. https://doi.org/10.6028/NIST.IR.8351

[NISTIR 8351sup1] Butler, J.M. (2024) History of DNA Mixture Interpretation. *Supplemental Document to DNA Mixture Interpretation: A NIST Scientific Foundation Review.* (National Institute of Standards and Technology, Gaithersburg, MD), NIST IR 8351sup1. https://doi.org/10.6028/NIST.IR.8351sup1

Noël S, Noël J, Granger D, Lefebvre JF, Séguin D (2019) STRmix™ put to the test: 300 000 non-contributor profiles compared to four-contributor DNA mixtures and the impact of replicates. *Forensic Science International: Genetics* 41:24-31. https://doi.org/10.1016/j.fsigen.2019.03.017

Oostdik K, Lenz K, Nye J, Schelling K, Yet D, Bruski S, Strong J, Buchanan C, Sutton J, Linner J, Frazier N, Young H, Matthies L, Sage A, Hahn J, Wells R, Williams N, Price M, Koehler J, Staples M, Swango KL, Hill C, Oyerly K, Duke W, Katzilierakis L, Ensenberger MG, Bourdeau JM, Sprecher CJ, Krenke B, Storts DR (2014) Developmental validation of the PowerPlex(®) Fusion System for analysis of casework and reference samples: A 24-locus multiplex for new database standards. *Forensic Science International: Genetics* 12:69-76. https://doi.org/10.1016/j.fsigen.2014.04.013

[PCAST 2016] President's Council of Advisors on Science and Technology. Report to the President - Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. September 2016. Available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf (accessed August 24, 2024).

Perlin MW & Sinelnikov A (2009) An information gap in DNA evidence interpretation. *PLoS ONE* 4(12):e8327. https://doi.org/10.1371/journal.pone.0008327

Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, Duceman BW (2011) Validating TrueAllele® DNA mixture interpretation. *Journal of Forensic Sciences* 56(6):1430-1447. https://doi.org/10.1111/j.1556-4029.2011.01859.x

Perlin MW, Belrose JL, Duceman BW (2013) New York State TrueAllele® casework validation study. *Journal of Forensic Sciences* 58(6):1458-1466. https://doi.org/10.1111/1556-4029.12223

Perlin MW, Dormer K, Hornyak J, Schiermeier-Wood L, Greenspoon S (2014) TrueAllele casework on Virginia DNA mixture evidence: computer and manual interpretation in 72 reported criminal cases. *PLoS ONE* 9(3):e92837. https://doi.org/10.1371/journal.pone.0092837

Perlin MW, Hornyak JM, Sugimoto G, Miller KW (2015) TrueAllele® genotype identification on DNA mixtures containing up to five unknown contributors. *Journal of Forensic Sciences* 60(4):857-868. https://doi.org/10.1111/1556-4029.12788

Perlin MW (2018) Efficient construction of match strength distributions for uncertain multi-locus genotypes. *Heliyon* 4(10):e00824. https://doi.org/10.1016/j.heliyon.2018.e00824

Prieto L, Haned H, Mosquera A, Crespillo M, Aleman M, Aler M, Alvarez F, Baeza-Richer C, Dominguez A, Doutremepuich C, Farfán MJ, Fenger-Grøn M, García-Ganivet JM, González-Moya E, Hombreiro L, Lareu MV, Martínez-Jarreta B, Merigioli S, Milans Del Bosch P, Morling N, Muñoz-Nieto M, Ortega-González E, Pedrosa S, Pérez R, Solís C, Yurrebaso I, Gill P (2014) Euroforgen-NoE collaborative exercise on LRmix to demonstrate standardization of the interpretation of complex DNA profiles. *Forensic Science International: Genetics* 9:47-54. https://doi.org/10.1016/j.fsigen.2013.10.011

Puch-Solis R, Rodgers L, Mazumder A, Pope S, Evett I, Curran J, Balding D (2013) Evaluating forensic DNA profiles using peak heights, allowing for multiple donors, allelic dropout and stutters. *Forensic Science International: Genetics* 7(5):555-563. https://doi.org/10.1016/j.fsigen.2013.05.009

Puch-Solis R & Clayton T (2014) Evidential evaluation of DNA profiles using a discrete statistical model implemented in the DNA LiRa software. *Forensic Science International: Genetics* 11:220-228. https://doi.org/10.1016/j.fsigen.2014.04.005

[QAS 2020] Federal Bureau of Investigation (2020) Quality Assurance Standards (QAS) for Forensic DNA Testing Laboratories. Available at https://www.swgdam.org/publications (accessed July 1, 2024).

Riman S, Iyer H, Vallone PM (2021) Examining performance and likelihood ratios for two likelihood ratio systems using the PROVEDIt dataset. *PLoS ONE* 16(9): e0256714. https://doi.org/10.1371/journal.pone.0256714

Riman S, Iyer H, Vallone PM (2022) A response to a correspondence letter by Buckleton et al. on: Riman et al. (2021) Examining performance and likelihood ratios for two likelihood ratio systems using the PROVEDIt dataset, PLoS One 16(9):e0256714. *Forensic Science International: Genetics* 59: 102710. https://doi.org/10.1016/j.fsigen.2022.102710

Riman S, Bright JA, Huffman K, Moreno LI, Liu S, Sathya A, Vallone PM (2024a) A collaborative study on the precision of the Markov chain Monte Carlo algorithms used for DNA profile interpretation. *Forensic Science International: Genetics* 72:103088. https://doi.org/10.1016/j.fsigen.2024.103088

Riman S, Iyer H, Robitaille E, Liu S, Sathya A, Vallone PM (2024b) A Software Package for Designing and Interpreting Forensic DNA Validation Studies. Poster by Sarah Riman at the International Society for Forensic Genetics on September 11, 2024. Available at

https://strbase.nist.gov/NIST_Resources/Presentations/2024/ISFG2024_Riman_SoftwareTool.pdf (accessed October 8, 2024).

Rodriguez JJRB, Bright J-A, Salvador JM, Laude RP, de Ungria MCA (2019) Probabilistic approaches to interpreting two-person DNA mixtures from post-coital specimens. *Forensic Science International:* 300:157-163. https://doi.org/10.1016/j.forsciint.2019.04.037

Rogers E, Aranda R, Spencer PM, Myers DR (2022) DNA mixture study: Novel metrics to quantify the intra- and inter-laboratory variability in forensic DNA mixture interpretation. National Institute of Justice report on award 2013-DNR-5042. Available at https://www.ojp.gov/pdffiles1/nij/grants/304317.pdf (accessed July 5, 2024).

Royal Society (2017) *Forensic DNA Analysis: A Primer for Courts*. Available at https://royalsociety.org/about-us/what-we-do/science-and-law/ (accessed October 26, 2024).

Russell L, Cooper S, Wivell R, Kerr Z, Taylor D, Buckleton J, Bright JA (2019) A guide to results and diagnostics within a STRmix™ report. *Wiley Interdisciplinary Reviews: Forensic Science* 1(6):e1354. https://doi.org/10.1002/wfs2.1354

[Sacramento County] ESR (2017) Internal Validation of STRmix v2.4 Sacramento County District Attorney's Crime Laboratory. Available at https://indefenseof.us/uploads/Sacramento-Cty-DA-STRmix-V2-4-internal-validation-summary.pdf (accessed October 21, 2024).

Schuerman C, Kalafut T, Buchanan C, Sutton J, Bright JA (2020) Using the nondonor distribution to improve communication and inform decision making for low LRs from minor contributors in mixed DNA profiles. *Journal of Forensic Sciences* 65(4):1072-1084. https://doi.org/10.1111/1556-4029.14306

Steele CD, Greenhalgh M, Balding DJ (2014) Verifying likelihoods for low template DNA profiles using multiple replicates. *Forensic Science International: Genetics* 13:82-89. https://doi.org/10.1016/j.fsigen.2014.06.018

Steele CD, Greenhalgh M, Balding DJ (2016) Evaluation of low-template DNA profiles using peak heights. *Statistical Applications in Genetics and Molecular Biology* 15(5): 431-445. https://doi.org/10.1515/sagmb-2016-0038

Susik M & Sbalzarini IF (2023) Analysis of the Hamiltonian Monte Carlo genotyping algorithm on PROVEDIt mixtures including a novel precision benchmark. *Forensic Science International: Genetics* 64:102840. https://doi.org/10.1016/j.fsigen.2023.102840

Swaminathan H, Garg A, Grgicak CM, Medard M, Lun DS (2016) CEESIt: A computational tool for the interpretation of STR mixtures. *Forensic Science International: Genetics* 22:149-160. https://doi.org/10.1016/j.fsigen.2016.02.005

Swaminathan H, Qureshi MO, Grgicak CM, Duffy K, Lun DS (2018) Four model variants within a continuous forensic DNA mixture interpretation framework: effects on evidential inference and reporting. *PLOS ONE* 13(11): e0207599. https://doi.org/10.1371/journal.pone.0207599

SWGDAM (2010) SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories.

SWGDAM (2015) SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems. Available at https://www.swgdam.org/publications (accessed July 1, 2024).

SWGDAM (2017) SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. Available at https://www.swgdam.org/publications (accessed August 24, 2024).

SWGDAM (2018) Recommendations of the SWGDAM Ad Hoc Working Group on Genotyping Results Reported as Likelihood Ratios. Available at https://www.swgdam.org/publications (accessed July 1, 2024).

Taylor D, Bright JA, Buckleton J (2013) The interpretation of single source and mixed DNA profiles. *Forensic Science International: Genetics* 7(5):516-528. https://doi.org/10.1016/j.fsigen.2013.05.011

Taylor D (2014) Using continuous DNA interpretation methods to revisit likelihood ratio behaviour. *Forensic Science International: Genetics* 11:144-153. https://doi.org/10.1016/j.fsigen.2014.03.008

Taylor D & Buckleton J (2015) Do low template DNA profiles have useful quantitative data? *Forensic Science International: Genetics* 16:13-16. https://doi.org/10.1016/j.fsigen.2014.11.001

Taylor D, Buckleton J, Evett I (2015) Testing likelihood ratios produced from complex DNA profiles. *Forensic Science International: Genetics* 16:165-171. https://doi.org/10.1016/j.fsigen.2015.01.008

Taylor D, Bright JA, McGovern C, Hefford C, Kalafut T, Buckleton J (2016a) Validating multiplexes for use in conjunction with modern interpretation strategies. *Forensic Science International: Genetics* 20:6-19. https://doi.org/10.1016/j.fsigen.2015.09.011

Taylor D, Buckleton J, Bright JA (2016b) Factors affecting peak height variability for short tandem repeat data. *Forensic Science International: Genetics* 21:126-133. https://doi.org/10.1016/j.fsigen.2015.12.009

Taylor D, Bright JA, Kelly H, Lin MH, Buckleton J (2017a) A fully continuous system of DNA profile evidence evaluation that can utilise STR profile data produced under different conditions within a single analysis. *Forensic Science International: Genetics* 31:149-154. https://doi.org/10.1016/j.fsigen.2017.09.002

Taylor D, Buckleton J, Bright JA (2017b) Does the use of probabilistic genotyping change the way we should view sub-threshold data? *Australian Journal of Forensic Sciences* 49(1):78-92. https://doi.org/10.1080/00450618.2015.1122082

Taylor D, Curran JM, Buckleton J (2017c) Importance sampling allows Hd true tests of highly discriminating DNA profiles. *Forensic Science International: Genetics* 27:74-81. https://doi.org/10.1016/j.fsigen.2016.12.004

Thomson J (2018) AFSP Mixture Evaluation Collaborative Exercise. Presentation given at the English Speaking Working Group of the International Society for Forensic Genetics held September 2018 in St.

Petersburg, Russia. Available at http://www.isfg.org/files/ESWG2018_Thomson_AFSP_Mixture.pdf (accessed August 8, 2024).

Toscanini U, Gusmao L, Alava Narvaez MC, Alvarez JC, Baldassarri L, Barbaro A, Berardi G, Betancor HE, Camargo M, Carreras-Carbonell J, Castro J, Costa SC, Coufalova P, Domínguez V, Fagundes de Carvalho E, Ferreira STG, Furfuro S, García O, Goios A, González R, de la Vega AG, Gorostiza A, Hernández A, Jiménez Moreno S, Lareu MV, León Almagro A, Marino M, Martínez G, Miozzo MC, Modesti NM, Onofri V, Pagano S, Pardo Arias B, Pedrosa S, Penacino GA, Pontes ML, Porto MJ, Puente-Prieto J, Pérez RR, Ribeiro T, Rodríguez Cardozo B, Rodríguez Lesmes YM, Sala A, Santiago B, Saragoni VG, Serrano A, Streitenberger ER, Torres Morales MA, Vannelli Rey SA, Velázquez Miranda M, Whittle MR, Fernández K, Salas A (2016) Analysis of uni and bi-parental markers in mixture samples: Lessons from the 22nd GHEP-ISFG Intercomparison Exercise. *Forensic Science International: Genetics* 25:63-72. https://doi.org/10.1016/j.fsigen.2016.07.010

[UKFSR 2018] UK Forensic Science Regulator (2018) DNA mixture interpretation, FSR-G-222. Available at https://www.gov.uk/government/publications/dna-mixture-interpretation-fsr-g-222 (accessed July 1, 2024).

Walsh PS, Erlich HA, Higuchi R (1992) Preferential PCR amplification of alleles: mechanisms and solutions. *PCR Methods & Applications* 1(4):241-250. https://doi.org/10.1101/gr.1.4.241

Wang DY, Chang CW, Lagacé RE, Calandro LM, Hennessy LK (2012) Developmental validation of the AmpFℓSTR® Identifiler® Plus PCR Amplification Kit: an established multiplex assay with improved performance. *Journal of Forensic Sciences* 57(2):453-465. https://doi.org/10.1111/j.1556-4029.2011.01963.x

Ward D, Henry J, Taylor D (2022) Analysis of mixed DNA profiles from the RapidHIT ID platform using probabilistic genotyping software STRmix. *Forensic Science International: Genetics* 58: 102664. https://doi.org/10.1016/j.fsigen.2022.102664

You Y & Balding D (2019) A comparison of software for the evaluation of complex DNA profiles. *Forensic Science International: Genetics* 40:114-119. https://doi.org/10.1016/j.fsigen.2019.02.014